LEVEL II

12

NOSC

Technical Report 380

REMOTE OPERATOR PERFORMANCE COMPARING MONO AND STEREO TV DISPLAYS: THE EFFECTS OF VISIBILITY, LEARNING AND TASK FACTORS

DC Smith
Naval Ocean Systems Center

RE Cole
University of Hawaii

JO Merritt
Human Factors Research, Inc.

RL Pepper
Naval Ocean Systems Center

February 1979

Prepared for
Engineering Psychology Programs
Office of Naval Research

Approved for public release; distribution unlimited

NAVAL OCEAN SYSTEMS CENTER
SAN DIEGO, CALIFORNIA 92152

NAVAL OCEAN SYSTEMS CENTER, SAN DIEGO, CA. 92152

AN ACTIVITY OF THE NAVAL MATERIAL COMMAND

RR GAVAZZI, CAPT, USN
Commander

HOWARD L BLOOD, PhD
Technical Director

ADMINISTRATIVE STATEMENT

The research described in this report was conducted at the Naval Ocean Systems Center, Hawaii Laboratory, and sponsored by the Office of Naval Research Psychological Sciences Division under work unit number 196-150.

The authors extend their appreciation to Mr. Gerald Malecki and Dr. Martin Tolcott of the Engineering Psychology Programs Office of ONR for their support during the course of this research.

We wish also to acknowledge the support of various people within the NOSC Laboratory who contributed significantly to the work reported here: Ms. Malia Nieves for her diligent work in running subjects and analyzing data, Carl Suekawa for his technical skills in maintaining much of the electronic equipment and Dick Brady and Patsy Kaneshiro for their technical editing work.

Released by
JAMES K. KATAYAMA, Head
Ocean Sciences Division

Under Authority of
J. D. HIGHTOWER, Head
Environmental Sciences Department

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: NOSC TR-380	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Remote Operator Performance Comparing Mono and Stereo TV Displays: the Effects of Visibility, Learning and Task Factors.		5. TYPE OF REPORT & PERIOD COVERED: Final Technical Rep[redacted]t
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): D. C. Smith, R. E. Cole, J. O. Merritt R. L. Pepper		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Ocean Systems Center, Hawaii Laboratory, P. O. Box 997, Kailua, Hawaii 96734		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61153/R04207; RR040901; 196-150.
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Engineering Psychology Programs, Code 455, 800 North Quincy St., Arlington, VA 22217		12. REPORT DATE: Feb[redacted]79
		13. NUMBER OF PAGES: 46
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 47p		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

stereo viewing; television; human perceptual performance.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An analysis of visual perception related to the underwater environment is presented, followed by a discussion of 3-dimensional space perception, and the absolute and relative cues employed by the operator during task performance.

(continued)

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

393 159

Task factors are described in simple terms and related to both real-world tasks and tasks which have evolved in the laboratory to test the real-world components. Arbitrarily designated as category 1, 2 and 3 tasks, these tasks are differentially loaded on at least 3 dimensions: Visual complexity, the magnitude of depth plane positioning required by the operator, and the requirement for scene interpretation by the operator.

In an effort to support the ideas generated by our analysis of visibility, task, and learning factors, three experiments were conducted.

Using a category 1 task, experiment 1 employed highly practiced subjects to reduce the effects of learning. Mono and stereo TV performance was measured under three levels of visibility degradation (simulated by contrast reduction). As predicted, stereo was superior to mono under all conditions tested. Performance using both mono and stereo displays were both affected by degraded visibility.

Experiment 2 was conducted with naive subjects using an experimental design which enabled an assessment of the degree of learning under operator testing conditions. We hypothesized that the category 1 task would show significantly less advantage for stereo, but that the effects of degraded visibility would continue to occur. The results are consistent with our interpretation.

In experiment 3, the more visually complex category 2 task was employed. The design of the experiment was similar to experiment 2 so that evidence for learning could be assessed under these different task conditions. Predictions concerning the degree of performance advantage for stereo vs mono displays were supported. This advantage was observed to increase with decreasing visibility, a finding which is consistent with our earlier predictions.

Conclusions and recommendations for further research aimed at understanding the relative contributions of several additional factors which operate to determine visual perception are discussed.

A final discussion of the need for further research in visual perception ends with recommendations for future investigation of the role of several additional factors (motion parallax and visual-motor space) in perception.

EXECUTIVE SUMMARY

The following report is a culmination of two years' work. This research was initially stimulated by the contradiction occurring in the literature that despite the large differences in performance under binocular and monocular direct-viewed testing conditions, comparable testing with mono and stereo TV showed little or no advantage for the stereo systems. This literature is briefly summarized, followed by a review of the preliminary research conducted in our laboratory.

An analysis of the problems involved in performance assessment with televised display systems led us to the conclusion that, in addition to the requirement of a well-organized, optically adequate and precisely calibrated stereo display system, visibility, task, and learning factors all act in combination to determine operator performance in comparison tests of TV display systems.

These three factors are described in the following sections.

TABLE OF CONTENTS

INTRODUCTION . . . 1

BACKGROUND . . . 1

LITERATURE REVIEW . . . 3

DISPLAY SYSTEM PERFORMANCE FACTORS . . . 4

VISIBILITY FACTORS . . . 4

VISUAL PERCEPTION IN 3-DIMENSIONAL SPACE: AN OVERVIEW . . . 5

Object Perception as Retinal Image Interpretation . . . 7
Classical Cue Theory . . . 10

TASK FACTORS . . . 14

LEARNING FACTORS . . . 15

TESTING CONDITIONS . . . 18

Manipulator . . . 18
Visibility Simulation . . . 18
Stereo Presentation . . . 20

LABORATORY EXPERIMENTS . . . 20

Peg-Task Experiments . . . 21
Experiment 1: Practiced Subjects . . . 21
Results and Discussion . . . 26
Experiment 2: Unpracticed Subjects . . . 29
Results and Conclusions . . . 30
Messenger-Line-Feeding (MLF) Task Experiment . . . 32
Experiment 3. . . 32
Results and Discussion . . . 34

CONCLUSIONS AND RECOMMENDATIONS . . . 36

REFERENCES . . . 39

FIGURES

1. Diagram of Retinal Image Concept in display system design.
2. Subjective contours clearly show the inferential nature of visual perception.
3. The inferential process of perception.
4. Non-stereo imagery which illustrates the inferential process of perception.
5. Patterns with strong non-stereo depth cues.
6. Learning effects of repeated testing on a manipulator positioning task.
7. Peg Task in Typical Board Position.
8. Elevation and rotation combinations for the Peg-Task.
9. Test subject position with manipulator and TV.
10. Example of monocular cue.
11. Example of Interposition Cue.
12. Closer view of TV display screen, showing some of the perspective cues to alignment.
13. TV monitor in stereo mode and in mono mode.
14. Some examples of geometric perspective cues.
15. Average Peg-Task performance times for six practiced subjects.
16. Examples of raster-line interference.
17. Average Peg-Task performance times for unpracticed subjects.
18. Average Peg-Task error scores for unpracticed subjects.
19. Overall view of cameras and MLF taskboard.
20. Average Messenger-Line-Feeding (MLF) task performance times.
21. Average Messenger-Line-Feeding (MLF) task error performance.

TABLES

1. Visual cues to depth.
2. Analysis of variances practiced Peg-Task (Log Time).
3. Analysis of variance unpracticed Peg-Task (Log Time).
4. Analysis of variance unpracticed Peg-Task (Errors).
5. Analysis of variance of MLF task (Log Time).
6. Analysis of variance of MLF task (Errors).

INTRODUCTION

BACKGROUND

With the advent of space exploration and the recently intensified work to obtain undersea resources, there has been a growing interest in the use of unmanned systems for reconnaissance and for performing work with remotely controlled manipulators. The increasing variety and sophistication of remotely manned systems has resulted in a renewed interest in stereoscopic television as a display technology for improving remote teleoperator performance. This report is directed toward determining the utility of stereo TV for remotely manned system visual displays, with particular emphasis on specific problems encountered in the undersea environment.

There is nothing new about stereo imagery; it was once the equivalent of television as an evening entertainment for the family, and hundreds of thousands of stereograms (a pair of stereo images) were photographed and circulated before the turn of the century. The modern-day descendent of the old stereoscope is the View-master device, popular with today's younger generation. Although stereo viewing equipment and devices such as binoculars are readily accepted and widely used, stereo movies taken during the 1950's and attempts to obtain stereo TV have given stereo a bad name with the public at large. Complaints of visual discomfort were, and unfortunately still are, common for users of many stereo viewing systems. But just as it is possible to produce the engineering precision necessary to make binoculars comfortable and acceptable for prolonged use, it is possible to design and maintain a stereo television system which provides the benefits of stereo without the previously all-too-common eyestrain. The main problems with stereo TV systems result from the methods that have been used to separate the image channels so that each eye sees only its proper half of the stereo pair. These methods, variously employing Fresnel lenses, mirrors, prisms, beamsplitters, crossed polarizers, lenticular screens, flickering shutter glasses, and other such components, may cause optical degradation or perceptual interference relative to the level of quality available in a conventional monoscopic system. New technology in imaging equipment is currently experiencing rapid development. In much the same way that pocket calculators and microcomputers now make it possible to do what was prohibitively expensive and complex in the recent past, it will soon be possible to employ simple, lightweight solid-state cameras and display monitors which completely eliminate the usually difficult problems of stereo image matching and registration.

The added realism and spatial orientation provided by a properly adjusted stereo TV is impressive; however, the question of performance advantages relative to non-stereo TV must be directly addressed: can the operator do as well or almost as well with a conventional mono TV display system? In order to answer this question, we must consider the problems encountered by the operator in performing various underwater tasks.

The operator of a remotely manned system typically uses television to position the platform or vehicle at the work site. He then employs the manipulator to conduct prescribed tasks such as turning a valve or drilling a hole: tasks whose major requirement is eye-hand-manipulator coordination. Many mono/stereo comparison studies have been conducted under excellent visibility conditions using familiar work objects and simple prescribed tasks. The results of such studies have been used to evaluate the merits of stereo TV displays relative to mono systems. Yet, remote vehicle operators report that such ideal conditions are seldom encountered in day-to-day operations and that such tasks represent only some of the broad range of perceptual problems that they face. For example, a very important aspect of remote viewing that is often overlooked is interpretability. Scene interpretation plays a very important role in approaching the work site and positioning the vehicle. This differs markedly from the prescribed tasks referred to above. It usually involves no eye-hand-manipulator coordination, offers little opportunity for learning, and is usually conducted under degraded visibility conditions with unfamiliar or camouflaged objects. Failure to correctly interpret the televised scene during these positioning maneuvers can lead to slower task performance and errors which could result in damage to costly equipment or vehicle entanglement. Stereo significantly reduces interpretation problems.

In this report we will test the hypothesis that stereo TV will provide significant performance advantages for the operator of an undersea, remotely manned vehicle under a number of conditions. From an analysis of previous research, the following advantages are suggested, and need to be examined empirically:

1) Reduced search time for locating target objects and work areas.

2) Increased accuracy and reduced time for positioning the vehicle; also, reduced disturbance of bottom sediment and the subsequent time spent waiting for the water to clear.

3) Reduction in time required to perform tasks in which the visual parameters are the main determiner of performance.

4) Reduced reliance on "contact feedback" which might damage the work object or place it in an awkward recovery or work position.

5) Increased accuracy of tool positioning and manipulation, with less possibility of dropping or damaging tools (i.e., drill breakage, cross threading, jamming, etc.).

These advantages are expected to increase when the task involves (a) degraded visibility conditions, (b) unfamiliar or obscure targets, (c) task conditions which require precise manipulator positioning without "contact" feedback, and (d) single operation tasks where trial and error is unavailable to provide immediate perceptual-motor learning.

It should not be surprising that stereo will provide these advantages, because the use of stereo viewing equipment is considered virtually essential for routine use in a variety of fields which share much in common with the remote control of manipulators and unmanned submersibles. Stereo microscopes are widely used for industrial assembly of small components such as integrated circuits; eye surgery is performed with the aid of stereoscopic operating

microscopes; ophthalmologists routinely use stereo photography to record the contours of the retina and optic disc, and use stereo slit-lamp equipment to examine the cornea and lens of the eye; micro-surgery of millimeter-size blood vessels requires stereo viewing equipment; stereo X-Ray techniques are used to study the circulatory system of the brain; photo-interpreters use stereo viewing equipment to enhance the detection and recognition of significant objects, especially when interpretability is poor due to camouflage, object complexity, low contrast, graininess, etc.; and as a last and most familiar example, the use of binoculars as opposed to monocular telescopes shows that stereo viewing equipment can be preferred and accepted by the vast majority of individuals when properly designed, constructed, and aligned.

The pessimistic picture which emerges from the literature review in the following section will suggest that there is little to be gained from the extra cost and complexity of stereo TV. We will contend that these negative results are due to the uncontrolled effects of visibility conditions, learning factors and task characteristics that are not realistically related to the operational undersea environment, as well as due to the possibility of poor stereo alignment and registration.

LITERATURE REVIEW

This section will briefly review the limited number of studies in which a comparison was made between task performance with stereo TV and performance with non-stereo (i.e., mono) TV, plus several studies which evaluated stereo TV without comparing it to mono TV. As background for the effects of video display parameters on visual performance, the excellent and extensive review by Biberman (1973) covers a host of electro-optical variables such as resolution, field of view, contrast, granularity, and signal-to-noise ratio. For an excellent background reference on undersea imaging systems, the handbook by Funk, Bryant, and Heckman (1972) provides all levels of analysis from system trade-off decisions down to camera beam current values. However, there is no comparable work on performance, i. e., operator utilization of engineering or equipment parameters.

In the first paper to be covered, Chubb (1964) noted an unexpected result in stereo-mono performance comparison in a previous study by Kama and DuMars (1964). They found no significant differences in task performance times between mono TV and stereo TV, and in fact, the performance times with mono TV were faster than with stereo. Reasoning that problems with the stereo system could be the only explanation for stereo performance which was poorer than mono, Chubb designed a simple experiment to compare mono and stereo performance with direct viewing in place of the TV system. The test subjects used a through-the-wall manipulator arm to perform a fairly simple peg-in-hole task while viewing directly with their unaided eye (mono) or eyes (stereo) through a hot-cell window (radiation lab shielding). A simple clinical eyepatch was used to produce the mono condition. In contrast to Kama and DuMars' televised result, Chubb found that both the mean and variance of performance times were significantly increased in mono, with average performance time 20 percent more in mono than in stereo. He concluded that the lower resolution of the stereo TV system used by Kama and DuMars may have defeated whatever stereo advantages should have been present. In a number of studies comparing stereo TV with mono TV, the stereo system may have suffered from poorer resolution and difficulties of image matching and alignment, thus confounding the desired mono/stereo comparison with misalignment and eyestrain factors.

The first assessment of stereo TV for underwater application was reported by Pescl (1967). Using two tasks common to undersea salvage operations, he compared performance between a stereo and a mono display. Pesch concluded that the advantages given by a stereo display is task dependent, related to the visual environment, and sensitive to practice effects.

Hudson and Culpit (1968) assessed mono-stereo performance in a series of size and distance judgments. Under the condition of their experiment, no stereo advantage was observed.

NASA interest in viewing systems was pronounced during the early 1970's, with many aerospace contractors working on a variety of video problems (Essex Corp., RCA Astronautics, Martin Marietta, MB-Associates, and Stanford Research Institute). A paper by Pepper and Cole (1978) reviews this literature in detail. Their summary of the literature indicates that there is no consistent performance advantage using stereo TV compared with mono TV. They argued that this result is unexpected, based on the logic that binocular visual perception performance must always be as good as, or better than, monocular visual performance.

Pepper, Merritt, Cole, and Smith (1978) reported the results of three studies designed to compare operator performance in a variety of video display situations. The first two studies involved perceptual judgment; the third was a perceptual-motor task requiring the operator to position the end-effector of a manipulator. The results of Study 1 indicate that stereo performance is superior to mono performance using either a field sequential or a Fresnel stereo display system in a 2-rod depth discrimination task. Study 2 indicates that stereo thresholds obtained with Julesz random dot stereograms did not differ when employing a Fresnel or a field sequential stereo display; furthermore, the televised stereo thresholds did not differ appreciably from those obtained under direct-viewed conditions. In Study 3, a mono TV system was compared with the field sequential stereo system in a task requiring perceptual-motor coordination. Subjects were required to position the end-effector of a direct linkage manipulator directly over a designated attachment loop and grasp the loop appropriately with the end-effector. Time and error scores were recorded. Results indicate that the stereo display provides a significant advantage in both time to complete the response and in the errors made in executing the end-effector closure.

In a discussion of the implications of these and other research studies, Pepper and Cole concluded that performance was a complicated result of at least three factors acting in combination. These factors are the visual environment, the task itself, and the effects of operator learning. It seems appropriate to review the substance of those arguments at this time.

DISPLAY SYSTEM PERFORMANCE FACTORS

VISIBILITY FACTORS

In an undersea environment, the visibility factors which affect performance are the result of both physical and perceptual influences.

a) The physical effect of particulate matter in the water column results in backscattering of light, (i.e., veiling luminance). Additionally, visual noise results from the particulate

matter, creating a loss in display system resolution. The settling of these particles produces a camouflage effect which obscures edge and contour details of objects.

b) The perceptual influence of veiling luminance results in a contrast reduction between an object of interest and the scene background. This in turn affects the visual discriminability of these objects. Visual noise reduces picture resolution, which in turn will affect detection, discrimination, and object recognition. The camouflage effects of sediment and growth make objects imperceptible, uninterpretable, or indistinguishable from the scene background.

VISUAL PERCEPTION IN THREE-DIMENSIONAL SPACE: AN OVERVIEW

In order to more fully appreciate the process by which visual information is transformed into object percepts by the human visual system, the following rather lengthy and detailed discussion has been developed (Merritt, 1978). It has been prepared especially to facilitate an understanding of the perceptual cue complexities involved in video display systems, with particular reference to an underwater environment.

The visual process of object perception may be separated into two distinctly different components: (1) perception of an object's shape and color/reflectance/surface-texture, and (2) perception of an object's distance along the line of sight (spatial localization in the third dimension). The object's shape in the two-dimensional X–Y plane perpendicular to the line of sight is essentially analogous to its optical projection in the retinal image (a "flat" two-dimensional surface), and thus the visual perception of shape has not seemed as paradoxical as the perception of depth or distance along the third-dimensional Z-axis. Since the three-dimensional array of objects in space is optically collapsed into a two-dimensional range, it is difficult to suggest a process which could reconstitute or recover this lost third-dimensional information; to say that depth, or Z-axis distance, is perceived because of "depth cues" in the retinal image is somewhat circular, but at present there is an active research effort to answer these questions (which have been central issues in psychology since the mid-1800's). The ongoing work in machine intelligence and pattern recognition has served to point out that even simple shape recognition cannot be easily explained; we simply do not know how the human (or animal) visual system actually processes the retinal image in order to arrive at object-percepts localized in space in front of the observer. It is beyond the scope of this report to explore these intriguing problems further, but it is sufficient to say that visual perception is somehow inferred from the retinal images on the two eyes (or one eye if that is the case) and from the adjustments of the muscles that point the eyes and focus the retinal images. For the purpose of display system research, then, it is enough to conclude that the ultimate in remote viewing systems would fully duplicate, in the observer's left and right eyes, the retinal images and the oculomotor adjustments which would exist if the observer were at the actual remote location using direct viewing. Since our visual perception system cannot go "out beyond" the retina, any display which provides retinal images identical to those produced in the usual way by real objects will inevitably cause us to perceive those images as if the objects were really there. The most common case in which this occurs is seeing objects "in" the space behind a high quality mirror: even though we know there are no objects where they appear to be, the retinal images are identical to what would be formed by objects seen through a transparent window rather than "in" a reflective mirror. This somewhat overstated discussion is to emphasize the overly simple but very important concept that the objective of any display

system is to produce some kind of retinal image, and the display engineer is free to accomplish this by any means which suits the requirements for image information transfer and practicality of equipment. This display concept is illustrated in Figure 1.

The ways in which a display system fails to duplicate the full-cue situation (exactly those retinal images and oculomotor adjustments which would exist in direct viewing at the work site) can give rise to loss of visual information which is critical to the completion of some tasks, but which may be of little or no consequence for other tasks. We will repeatedly emphasize that a certain visual cue such as stereopsis (from binocular parallax disparity) may be very important for some tasks and of little importance for other tasks. This helps to explain the widely varying results in performance tests comparing stereo TV with conventional non-stereo TV.

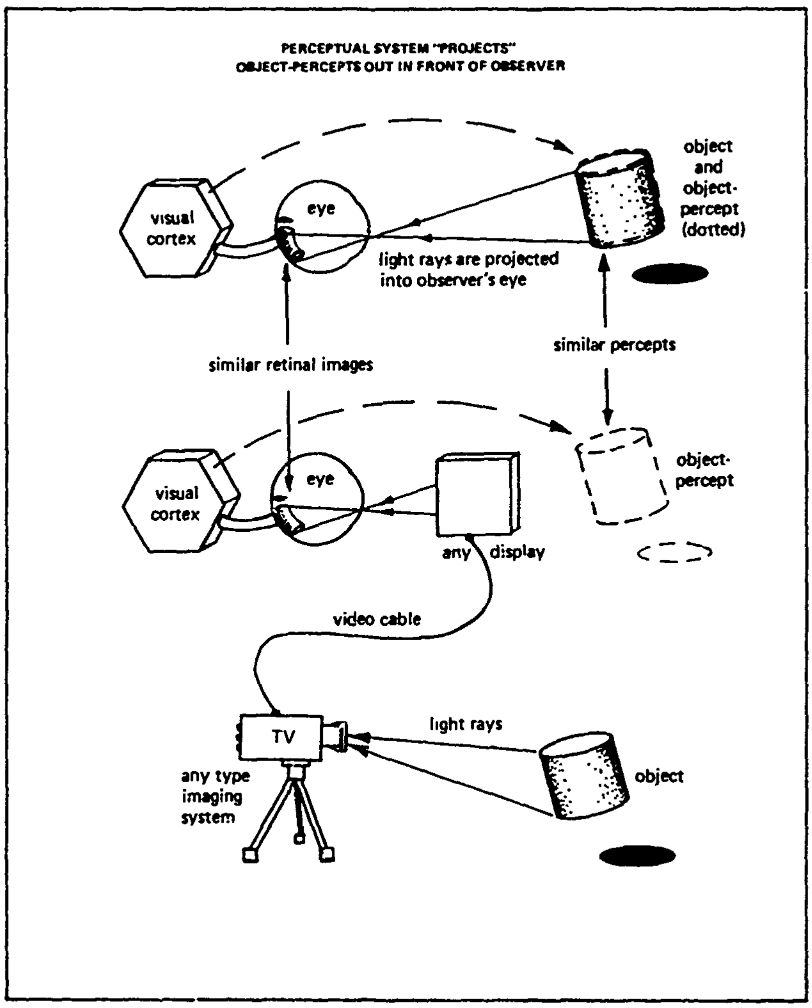


Figure 1. Diagram of Retinal Image Concept in display system design. All that is necessary for successful remote viewing systems is that ultimately the display must create the retinal images which would exist in the observer's left and right eyes if actually viewing the object directly, as at the top of the figure.

Object Perception as Retinal Image Interpretation

Everyday visual perception appears so veridical and rapid that we routinely assume that what we seem to see is in fact really there. We only reluctantly, and with great effort, accept the idea that what we see is not really the object itself; instead, what we see is the end result of a process of visual inference that goes on below the conscious level. In this modern age of computer "image understanding" systems, it would be appropriate to say that what we see is the "output display in graphic format" of the visual system's non-verbal interpretation report, showing what is most probably out there causing the current retinal imagery.

From the two small optical images, the perceptual system infers what is likely to be the cause of the retinal stimulation; these "visual inferences" are then displayed as "perceived objects in space," localized in front of the observer. The apparent spatial position of these object percepts represents the non-verbal way in which the perceptual system indicates its best guess about object size and distance.

Since it is only these perceptual object-inferences which are "seen," and never the objects themselves, the way is open for creating a display or simulator system which produces the appearance of objects when none are actually present. The computer-generated world produced in the increasingly realistic flight simulators is a good example; there, the perceived objects do not exist at all, even at a remote location.

The brain, working only with retinal images, has no more direct contact with the imaged objects than does a photointerpreter working with photographs of places he has never visited. The inferential process (at an unconscious level) is in many ways analogous to the process of photointerpretation at a conscious level; even the direction of eye fixations is analogous–when the lower-resolution peripheral retina detects something which warrants a better look, the oculomotor system orders a high-resolution photo coverage by pointing the fine-grained central retina to image the object. The inferential nature of the visual perception process is clearly visible in the phenomenon of "subjective" contours; in Figure 2, an intervening obscuring surface is inferred as the best reason for interruption of a most likely simple square object and a set of four probable full discs. For some reason, these inferred obscuring

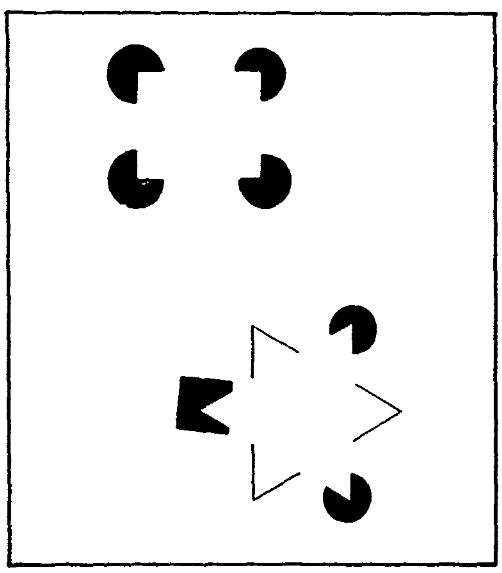


Figure 2. Subjective contours clearly show the inferential nature of visual perception. Note that the inference of a simpler, more probable, geometric shape requires the corollary inference of an "invisible" obscuring shape which is, interestingly, "whiter than white" in appearance. In a sense, all contours of perceived objects are "subjective" but are usually coincident with physical demarcations of luminance or hue.

surfaces appear to be "whiter than white" (Hennessy, 1975) because the visual system has no other way to differentiate the inferred object from the white background. It is difficult to say why all visually perceived contours are not equally "subjective." These subjective contours are admittedly different from the usual case in which the edge of an objectimage is demarcated by a change in luminance or some other physically measurable attribute. (Here, and for the remainder of this report, it should be kept in mind that many of these perceptual issues represent long-standing research questions; the simple characterization offered here is for the practical purpose of discussing problems in visual display of remotely manned manipulator operations. The vision research literature is teeming with alternative hypotheses regarding many aspects of visual perception.)

The process of object perception is a paradoxical one, as is admirably explained by Gregory (1966, 1970). As noted previously, due to the two-dimensional nature of the retinal image the three-dimensional information of objects-in-space is lost. This spatial information has to be reconstituted somehow by inference (not conscious inference, of course). This is paradoxical because a two-dimensional point on the retina represents a direction along a line of sight, but does not directly encode the distance along that line of sight. A set of shapes in the retinal image could result from an infinite number of real three-dimensional objects; somehow the visual system is able to tell which probable object is most likely, and choose that alternative for display. The visual system has to choose, on the basis of incomplete evidence, one of the possible objects which could have produced the retinal image. Since there is no algorithm we know of which can do this, the choice is based in some way on what is most likely to be found in the world of familiar things. This choice, or tendency to choose what is most probable, is seen in Figure 3, where although the only objects present are arrays

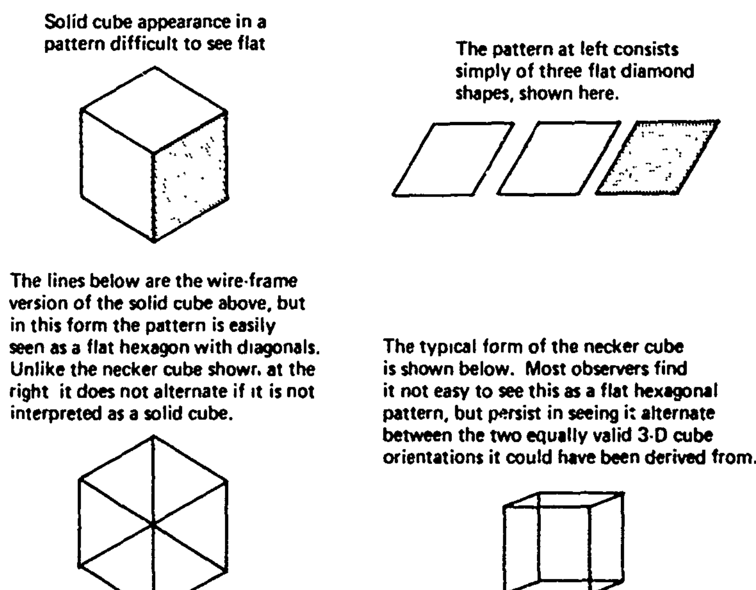


Figure 3. The inferential process of perception whereby the two-dimensional (flat) set of three diamond shapes in the upper right are almost involuntarily seen as a three-dimensional cube when arranged so as to produce a cube's retinal image, upper left. The job of the visual system could be characterized as making the best guess about what set of three-dimensional objects could be "out there" producing the current retinal images.

of black lines on the two-dimensional surface of the page, there is an involuntary and stubborn tendency to see a three-dimensional cube in upper left and lower right. In the lower right, the familiar phenomenon of alternating object-percepts is seen, where either of two (or more) objects could cause the retinal image. (Interestingly, the third interpretation as a flat pattern of rectangles, parallelograms, and triangles is hardly ever noticed.) Apparently, the solid cube is the more likely object to be expected in everyday experience.

The perceptual task of separating objects from their backgrounds (figure/ground problem) is easier when binocular parallax (stereo disparity) is available and the observer is not restricted to viewing a flat image of the scene. This provides the visual system with unambiguous primary depth cues which separate and delineate objects even before they are recognized by two-dimensional shape. (The ingenious random-dot stereograms presented by Julesz in 1971 illustrate this point.)

Without binocular parallax or motion parallax to help separate the jumbled 2–D object-images on the retina, the visual system seems faced with the circular paradox of having to first identify an object in order to pick it out from the background, but on the other hand having to pick it out in order to identify it by shape. Although the visual system does this routinely, no one has offered a satisfactory account of how it occurs. Figure 4 illustrates this fundamental problem in perception: the image is a flat pattern of light and dark, but it is also the 2–D projection of a familiar 3–D object. For most observers seeing this without prior knowledge or exposure, the retinal image goes uninterpreted, and the 2–D raw data on the retina is all that is perceived. It is as if the visual system accepts the 2–D raw data when it is unable to find a reasonable 3-D projection. This photograph is remarkable in that it slows down the process of perception so that we observe the process which usually occurs immediately. For most observers, the object percept of a white-faced calf with black ears forms suddenly upon hearing what hypothesis would give a good fit to the retinal facts. The figure also illustrates the phenomenal power of image memory when the reader views this photograph months or even years later and still sees the calf immediately. The calf, once seen, cannot be unseen, even when image quality is degraded still further; this image-memory capability is one of the factors which make proper performance evaluation of display systems subject to order effects.

Figure 4. Non-stereo imagery which illustrates the inferential process of visual perception. The visual system must make a best guess about the objects which could have caused this retinal image. Study the image first, then look again and see the white-faced calf with two black ears, looking straight at you. If this were displayed in stereo, there would be no such delay in perception.

It is important to note that there would have been no delay in perceiving the calf if stereoscopic photography had been used. Just as with a random-dot stereogram, the camouflaged image would stand out immediately without first having to be seen as a monocular contour. This consideration leads to the experimental hypothesis that stereo TV shows the greatest advantage over mono TV in those conditions where visibility and display factors degrade or eliminate the usual monocular cues to shape and distance. Thus, stereo also provides an interpretive function which is distinct from depth information given by retinal disparity.

Classical Cue Theory

In t' , section, we will consider the stimulus conditions which give rise to depth perception for the purpose of comparing various viewing systems with the full-cue situation inherent in direct viewing.

The reader with a background in visual perception may wish to skip this simplified, classical cue exposition and continue with task and learning factors involved in display research. It is presented here to anticipate misunderstandings which may occur in subsequent discussion of our research findings.

Traditionally, the cues to space or depth (distance along the line of sight) perception have been segregated into two kinds: those that require use of two eyes and those that require only one eye. These binocular and monocular cues can further be characterized as optical image cues or eye-muscle feedback cues. The binocular/monocular cues to depth are shown in Table 1.

Table 1. Visual Cues to Depth.

Binocular	• Convergence • Binocular Retinal Parallax
Monocular	• Accommodation • Motion Parallax • Perspective • Size of Familiar Objects • Light and Shadow • Interposition • Haziness of Distant Objects

There are several excellent accounts of classical cues to depth perception which will supplement the limited scope of this discussion. Among those which can be highly recommended are Graham (1965), Hochberg (1971), Forgus (1966), Ogle (1962), and Gregory (1966). One paper discusses cues to depth in the context of designing 3–D displays for various purposes (Vlahos, 1965). This is an excellent article regarding depth perception, and he makes the seldom-appreciated point that a 3–D display does not necessarily imply one based on binocular parallax: if the non-binocular cues to depth are strong enough, a robust 3–D percept will be created with "monocular" cues.

In the discussion of cues which follows, the point should be made that depth perception is the result of the complex interaction among the whole constellation of cues present in any given situation, and that it is not reasonable to predict the perceptual resultant by an analytic additive approach. It is not sufficient to specify the factors in terms of their isolated effects; instead, the empirical approach of trying the cues in various combinations, using the actual viewing situation, should be explored.

Cues to perception are necessary, but not always sufficient, for the occurrence of a visual percept. The word "cue" itself suggests the nature of the way visual cue content is used by the visual system: the cue must be there for a percept to occur, but the visual system may "miss the cue," so to speak, and the cue is not relevant. These points could be summarized as (1) cue threshold (minimum level required), (2) cue effectiveness (in a given multiple cue situation), and (3) cue relevance (to a given type of visual task)

The depth cues listed in Table 1 will be described briefly in the following paragraphs.

Convergence. Convergence is the amount of inward eye rotation which results in the interaction of the lines of sight. The degree of inward rotation provides a crude sense of absolute distance (near/far), and a relative sense of distance (between objects).This cue probably originates in sensing the neural commands given to the eye muscles (rather than coming from feedback sensors for eye position after the muscles act). Convergence is important for scaling the amount of depth which is created from a given amount of retinal disparity. This disparity-scaling mechanism must be considered when attempting to make a stereo display which appears linear in X, Y, and Z axes. This depth-constance system is apparently designed to compensate for the fact that retinal disparity falls off with the square of the distance, while linear size falls off directly with the distance, creating a Z to X–Y mismatch without the disparity scaling from convergence feedback. Convergence is one of the so-called primary cues to depth, inasmuch as it does not depend on interpretation of the image content.

Binocular Retinal Parallax. This cue is the one most often considered as the primary stimulus giving rise to a true space perception. The visual system is exquisitely sensitive to very small amounts of difference between the two eyes' retinal images; the stereo disparity thresholds measured in laboratory work have been as small as 10 seconds of arc, similar to the thresholds for vernier acuity. The degree of sensitivity suggests that stereo must have been very important at one point in man's development, even though in the geometrically predictable city environment, a one-eyed man can do very well. Stereo vision is almost essential for walking quickly through uneven ground in the woods, or for jumping from rock to rock down a mountain trail. To some extent, motion parallax can help, but for slower moving vehicles underwater, stereo provides disparity even when not moving.

Accommodation. Accommodation is the change in shape of the lens enabling it to focus a sharp image on the retina. It can be shown that the act of focusing can alter the perceived distance and size of an object-percept, even though focusing has little or no effect on the optical size of an image on the retina (if kept sharp by an artificial pupil). Although this, too, is a primary cue, it is relatively weak and limited to relatively close-in distances. There is an automatic link between accommodation and convergence, so that the eyes tend to focus at the distance where lines of sight are converged, and vice-versa. The accommodation and convergence cues are what could be called "anti-cues" (Vlahos 1965) when viewing a flat

2-D display, since they, along with binocular disparity, tend to suggest that there is only a flat picture-pattern rather than solid objects in space. This lack of change in focus and changing convergence can be anti-cues to realism and harmony among the other cues.

Motion Parallax. Motion parallax refers to the perception of object movement resulting from the observer translating his head. The magnitude and direction of movement is determined by the distance of the object from the fixation point. Thus, motion parallax is another primary cue available when the camera position can be translated laterally (rather than just panned from the same point). A sensor mounted on a moving vehicle has available a robust cue to distance in the velocity vectors present in the near and far field. No remote viewing systems utilize translation movement of the camera, and thus this powerful primary cue to depth is unrealized.

Perspective. The laws of geometric optics describe how image size is proportional to object distance. This results in development of distance cues from the decreasing size of similar objects, the linear-perspective convergence of roads and railroad tracks in the distance, the increasing density of texture gradients, the loss of resolution with distance, and the increasing height on the picture plane for farther objects.

Size of Familiar Objects. Given a familiar object of known objective size, the distance to it can be estimated by the size of its image on the retina (in terms of visual angle). This cue can have a powerful effect, given the presence of known-sized objects (such as a telephone pole or a familiar person). This requires image interpretation, so it would be classed as a secondary or derived cue.

Light and Shadow. A light source casting shadows from a direction other than along the camera line of sight provides a projection onto the level surface next to the object. In addition, lights can give a sense of solidity to objects by proper shadow modeling and shading on the object itself. Shadow cues are especially helpful to remote manipulator operators for determining when the arm is about to contact the bottom or some target object near the bottom.

Interposition. This cue is a very important one for determining relative depth in rank order (not in absolute or continuous-relative ways). If object A obscures object B, then A is closer than B, and so on. Complex arrays of objects can be rank-ordered in depth provided there is enough contrast between object reflectances to determine which is in front.

Haziness of Distant Objects. This primary cue is, like the texture gradient, somewhat independent of image interpretation. As distance increases, more and more air mass or water volume intervenes between camera and target, thus adding more veiling scattered light to the image, washing out contrast as a function of distance. It is easy to see that adding the same veiling light to both sides of the contrast ratio will dramatically reduce contrast: a 10:1 contrast ratio becomes 10+20:1+20, or 30:21, a poorer contrast by far, but nevertheless a good cue, especially underwater, where contrast falls off rapidly within short distances.

One of the main objectives of this report is directed toward comparing and contrasting performance on different tasks either with stereo TV or with conventional non-stereo TV. It is obvious that in a full-cue viewing situation, where there is a rich and redundant set of cues indicating object distances and identities, it would be possible to take away several

redundant cues without losing good depth perception. Those tasks which inherently have strong nonbinocular cues to depth may be performed almost as quickly without stereo as with it. Other tasks are virtually impossible without stereo, due to a lack of adequate depth and distance information. The following paragraphs describe additional ways that depth information can be used by remote underwater manipulator operators.

Line of Sight. Although this is not a depth cue, there is a technique developed by remote manipulator operators for working without a good sense of object distance. By superimposing the manipulator jaws (or end effector of any type) on the desired object while viewing the TV screen the operator simply keeps moving the manipulator along that line of sight until it contacts the object. This technique comes into play very often when working in non-stereo TV situations, and is one of the factors which makes for little mono-stereo performance differences in some types of tasks. Stereo permits approach to a target from directions other than along the line of sight. This is important when (1) travel along the line of sight is blocked by an obstruction or hazard, or (2) when the manipulator would block continuous visual contact with the desired object.

Gray Scale and Interposition Cues. Interposition cues may be severely degraded or absent in high-contrast imaging situations where a number of targets exceed the dynamic range of the TV system so as to appear all black or all white, thus giving no indication where they overlap. Similarly, when contrast is reduced by underwater visibility conditions, tonal shades of gray may be lost at the intersections between objects. Under such conditions, however, stereo (binocular disparity) would continue to provide sensitive and precise depth information after mono cues have been lost.

Resolution and Mono Cues. Certain mono cues to depth require significantly more resolution than do the stereo cues. Thus, a lower resolution stereo system which permits a wider field of vision can often deliver performance equal to a higher resolution mono system. This wider field of view could then in turn make certain types of tasks easier (e.g., keeping a sense of orientation to the sea floor and the work objects). Of course, there are some tasks for which high resolution is essential, with or without stereo, but in a majority of task situations, stereo can provide the same spatial response with less resolution than mono.

Mono-Stereo Cue Conflict. The relative strength of mono cues to shape and contour, even when pitted against good anti-cue information from stereo disparity, can be seen in Figure 5. Despite the sensitivity of stereo acuity which indicates a flat surface, the probability

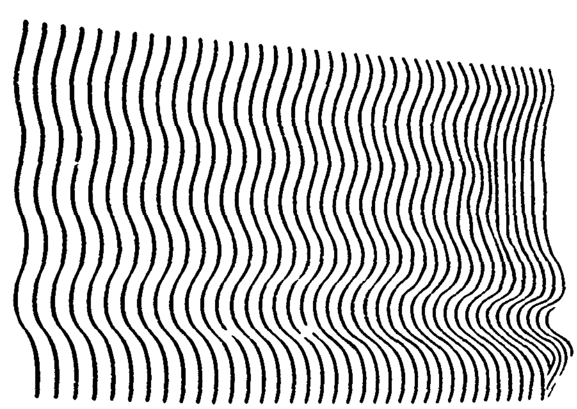


Figure 5. Patterns with strong non-stereo depth cues can overcome the anti-cue of stereo, which shows that the photograph is really flat on the page. The reader may see the effect as if the wavy portion is actually warped and curved.

of a curved surface, given the wavy lines, overrides the cues to flatness. For some viewers, the curvature cue is so strong that there is the suspicion that the paper has actually warped at the apparent ripple surface, and they can feel the sensory discrepancy by passing a fingertip over the figure.

Another familiar case in which the usually dominant cues from binocular disparity are defeated is illustrated by Gregory (1970), who shows that a human face presented in reversed stereo depth will not really look like the inside of a mask. Gregory notes, also, that the Necker cube (drawn in Figure 3) will still alternate in orientation when made into a wire-frame model (coated with phosphorescent paint so it will glow in the dark) and held in the hand of the observer. The completely unambiguous tactile information about the wire-cube's orientation is not enough to keep it from reversing!

The point of the previous discussion has been to show that binocular stereopsis is not the only true and powerful cue to depth; the ways in which it can be overcome by other cues and knowledge of the target point out the complexity of interaction among various cues.

The complex nature of the perceptual process of identification and localization make it difficult to reach definitive conclusions regarding the separate influence of a particular visual cue in the total interaction. At a different level of analysis, the relative contribution of the final perceptual process will be interwoven with the characteristics of those tasks which the operator is called upon to perform. We turn now to these task issues.

TASK FACTORS

Previous analysis of task factors led us to conclude that for the practical considerations of our research, most applied undersea manipulator tasks could be classified into three general categories based on similarities of their major perceptual-motor constituents.

Category 1 Tasks

(a) real world examples: drilling, tapping, threading, stacking, coupling, connecting.

(b) common components: alignment in the X (horizontal) and Y (vertical), little Z dimension positioning, frequent rotational movement.

(c) laboratory task: Peg-in-hole task as described by Hill and modified in our laboratory (Pepper and Cole, 1978).

Category 2 Tasks

(a) real world examples: line feeding, simple grabber attachment, sample recovery.

(b) common components: careful alignment in the X, Y and Z dimensions is required but the potential conflict with interposed elements between the object of interest and the camera system is reduced. Rich, visual scene with many conflicting objects.

(c) laboratory task: A messenger-line-feeding (MLF) task has been developed and tested. It is an elaboration of the end-effector positioning task employed by Pepper *et al.* (1978).

Category 3 Tasks

(a) real world examples: cable cutting, hooking and clamp attachments, flight recorder recovery.

(b) common components: precise alignment in the X and Y dimensions, greater need of positioning end-effector on the Z dimension, complex visual scene characterized by high degree of similar visual elements leading to confusion and interference from elements interposed between the object of interest and the camera system. Highly complex and ambiguous scene, with interpretation and recognition of objects required.

(c) No laboratory task yet developed, although complicated scenes and simulated flight recovery scenarios have been demonstrated.

LEARNING FACTORS

There are few situations when learning does not occur. Experiments which show learning effec s (when the primary concern is to evaluate performance effects) are the rule, rather than the exception. Learning occurs in both simple and complex tasks. The more complicated the task situation, the greater will be the learning effect. Also, the more complicated the task, the more complicated will be the analysis necessary to understand the relations between the learning effects and the contribution of task and visibility factors.

Learning is a pervasive phenomenon which occurs under both the real world conditions encountered by remote vehicle operators, as well as under laboratory conditions developed to test various components of these systems, including TV displays. In the underwater world, many tasks require repetition or successive approximation simply because "trial and error" may be the final, irreducible strategy available to the operator. While trial and error learning may l an essential part of the operator's strategy, one m.st recognize that it can be extremely costly either in operating time, or in increasingly risky or unsafe operating conditions. Any characteristic of a remotely operated system which speeds up learning, including enhancement of the information available to the operator through the image display system and proprioceptive feedback from the manipulator, will almost certainly result in a reduction in operating time, operating costs, and exposure to potentially hazardous situations.

While learning is important in the real world, it is in the laboratory that even greater concern for this phenomenon is required. This concern is necessitated by the frequent use of repeated trial designs which can quite easily confound learning with the effects of other independent variables. For example, Uhrich and Fugitt (1978), in testing two types of manipulator control and three viewing conditions, ran all subjects under all conditions and in the same order, yet they make no mention of possible learning effects in their interpretation.

Many of the researchers who attempt to account for the phenomena of learning treat it as a variable whose effects should be eliminated rather than studied for their practical and

theoretical consequences. Pesch (1967), for example, reports a mono-stereo difference that "washed out" on the second day's testing, implying that it was an unstable phenomenon of minor significance. In fact, the savings attributed to stereo might be very worthwhile, especially when we consider the improbability that a remote undersea vehicle operator performing a real life task would have two days of practice under precisely the same task and visibility conditions.

Another point to be made about learning phenomena has to do with their logical derivation from performance measures. As was mentioned before, repeated trial designs are often used in order to increase reliability of performance measures. It is important to note that the effects on performance that carry over from one trial to the next (called order effects) are the result of a complex interaction of a number of variables in addition to learning, including motivation, forgetting, and fatigue. Thus, performance levels can easily be misinterpreted. A case in point occurs when no improvement in performance occurs across a series of trials and is interpreted as an evidence of no learning effect. It is quite possible, especially in the case of manipulator tasks that require a good deal of physical force and movement, that increments in performance due to learning are cancelled out by the decremental effects of fatigue. A pilot study we conducted in developing our messenger-line-feeding task has bearing on this issue. A naive subject was given 30 trials a day, half mono and half stereo, for 10 days. For analysis of order effects within sessions, the fifteen trials for each viewing condition were divided into first five, second five, and third five trials. Results showed no improvement in performance within sessions for either mono or stereo viewing. A marked reduction in time scores did occur between sessions, however, as can be seen for the five sessions plotted in Figure 6. This result suggests that the subject was learning during a session but its effect on performance was counterbalanced by the decremental effects of fatigue. The obvious point here is that appropriate control conditions must be included in the design of an experiment in order to ensure clear interpretation of learning effects.

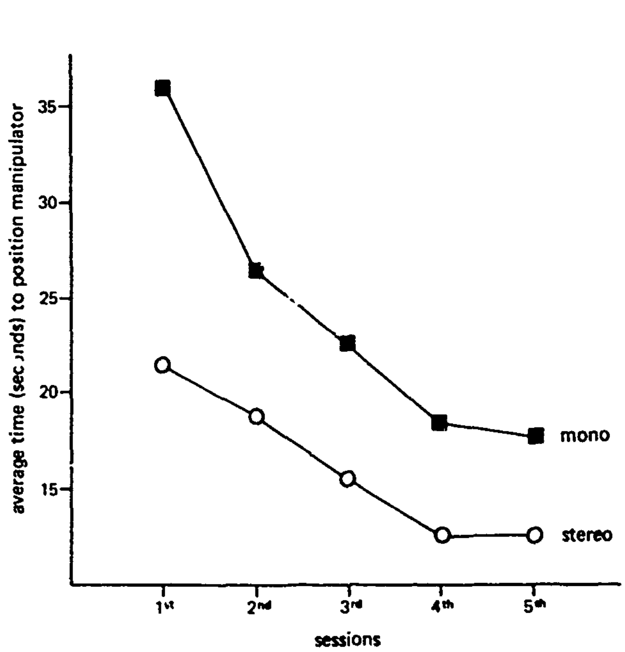


Figure 6. Learning effects of repeated testing on a manipulator positioning task.

A question that has important implications for the interpretation of learning effects is, "What does the operator learn?" For the completely naive subject on first entering the laboratory, there are a myriad of details to learn, including instructions, familiarity with the manipulator, task board, visual display, and procedures. Such learning-to-learn factors are present in all manipulator experiments and are usually accommodated by practice trials and coaching in the initial session and warm-up trials in the following testing sessions. However, despite these accommodations, considerable improvement in performance often occurs beyond the learning-to-learn stage, as is illustrated by our pilot subject's continued improvement in performance over many sessions.

At least two different types of learning would appear to determine performance on remote manipulator tasks: visual perceptual learning and motor learning. While these are not unrelated, it is likely that they are differentially affected by other performance variables, such as instructions, practice, fatigue, etc. An extensive analysis of the visual perceptual cues is contained in an earlier portion of this paper. It is sufficient to point out here that the visual scene may vary in complexity all the way from direct views of a simple, highly structured, totally familiar task board and manipulator arm to a barely discernible, complex TV display of an unfamiliar scene. The identification of critical features (form, shape, texture, etc.) of the task board and the location of objects in space constitute the two major components of the perceptual learning task. The rate at which these are learned will depend on the strength of the visual cues present, either in the scene itself, in the case of direct view, or on the monitor display, in the case of televised images. This state of affairs has some important implications for the choice of control conditions employed in studies designed to test the effects of variables on the rate at which visual, as opposed to motor, learning occurs. These will be discussed in a later section of this report.

The complexity of the motor learning requirements may also vary widely. The simple finger movements required by the switch closure apparatus employed by Uhrich and Fugitt (1978) and the restricted hand movements required by their joy stick manipulator represent relatively simple motor learning tasks. On the other hand, the directly linked, remote arm employed by Pepper *et al.* (1978) requires large coordinated movements of the upper body and arm, complex shoulder, elbow and wrist maneuvers, and hand closures. On the other hand, Pepper's manipulator provides force feedback on contact with objects, as well as arm-hand-body position cues that can be associated with the visual view of the remote arm relative to critical task board features. Thus, there is a rich assortment of motor cues available to aid perceptual motor learning with some manipulators. These cues may be greatly reduced or totally lacking in other experimental settings, which use other types of manipulators and controllers.

A number of conclusions and implications have resulted from our consideration of the role of learning factors in remote undersea manipulator problems.

1. Learning paradigms require proper control conditions in order that performance changes can be attributed to learning factors rather than other order effects.

2. Related but different kinds of learning may take place depending on task conditions, visibility conditions and the subject's experience.

Learning is nearly always present in both real world and laboratory situations. In interactions with task and visibility factors, it adds greatly to the problems of interpretation and generalization of research results. It is imperative, therefore, that we study learning effects with the same intensity and care given for other factors, rather than simply "control it out" of our research designs.

TESTING CONDITIONS

Before discussing the laboratory experiments a brief description of the manipulator, method of achieving reduced visibility, and type of stereo presentation is in order.

Manipulator

In all tests a standard Model G master-slave manipulator, built by Central Research Laboratories, was used. This direct linkage manipulator was designed to reproduce the natural movements and forces of the human hand at a remote location, i.e., an adjacent room or work location. The operator usually observes the end effector on the slave arm of the manipulator through a protective window, periscope, or as in our experiments, a television monitor. Except for slight amounts of deflection and the resulting lost motion, the manipulator end effector moves exactly as the operator moves the manipulator handle, no matter how complex the task motion may be, so long as it is within the dimensional limits of the manipulator. The forces at the end effector are equal to those applied by the operator at the handle, except for very slight amounts of friction and inbalance. This manipulator was chosen for our laboratory work because it is representative of the type of force feedback manipulators that will be available for undersea work systems in the future.

Visibility Simulation

As stated earlier, the main contributor to reduced underwater visibility is the backscatter of light from particulate matter suspended in the water column. In coastal waters the particulate matter is always present, while deep ocean water is clear and reduced visibility results when bottom sediment is stirred up by the undersea vehicle or work system.

In order to investigate operator performance under different levels of visibility, a procedure was developed to simulate backscatter (veiling luminance) in the laboratory. This procedure enabled the experimenter to present various levels of visibility to the operator during trial sequences.

The properties of closed-circuit TV systems make the problem of specifying visibility different from the usual optical measurement paradigm. The TV operator can compensate for a low contrast image at the camera faceplate by adjusting gamma or gain in the camera, or by adjusting the brightness and contrast at the monitor. This permits expansion of a light gray and dark gray into full black and white with a contrast transfer better than 100 percent at the monitor screen. There is a limit to this type of contrast enhancement, however, and when a given camera/monitor system has reached its limit, a gray and washed-out image may be the best

an operator has to work with. The various combinations of TV, monitor, lighting and water properties will result in a different quality of TV image presented to the operator, and thus a given screen image quality cannot be linked to a particular water property, i.e., an attenuation or scattering coefficient. What is important in the final analysis is the image delivered to the operator. It is this image which was experimentally varied.

The image on the TV monitor was measured in terms of the luminance (β) of the imaged reproduction of a known target placed in front of the cameras. Specifications for setting up the proper brightness and contrast on the TV monitor insured that all subjects receive the same visual input for each of the conditions.

The most appropriate way to relate levels of visibility used in our research to underwater optics is through the method of modulation transfer function (MTF) analysis. We assume the MTF of any remote TV viewing system is equivalent to that used in our laboratory. When a remote system in the real world encounters water conditions which interact with its imaging system to produce a particular quality of image on the monitor, then operator performance can be predicted by the MTF of the monitor image. See Funk, Bryant and Heckman (1972) for an appreciation of the factors affecting the monitor characteristics. Backscatter is the primary degrading factor in most remote system operations in the underwater environment, and is even more exaggerated in those systems that use their own illumination sources. It is fairly easy to simulate and measure backscatter, since the MTF of veiling luminance is simply a straight line showing equal contrast reduction for all spatial frequencies, regardless of the fineness of detail or the size of a dark area. Mertens (1970) provides an excellent and extensive treatment of the various component MTFs which cascade to produce the final overall system MTF in the underwater imaging situation. Since backscatter causes a veiling luminance which reduces contrast of both large and small details equally, it was controlled by means of the camera/monitor controls for brightness and contrast.

In order to present three different levels of visibility to the operator, a switching box was added to a Conrac QQA-17 black and white TV monitor. This modification enabled the contrast and brightness controls of the monitor to be paralleled by two other brightness and contrast controls. During alignment and calibration each of the three sets of controls were adjusted by the experimenter for different contrast ratios (visibility levels). In switch position one, the lighting, cameras and monitor were adjusted for the best overall presentation of the manipulator area. When this was determined, a test pattern with a white, gray and black area was placed in front of the cameras and the contrast ratio was determined using a Textronix Model J6523 Luminance meter. This became the baseline data for all further calibration tests. Positions two and three of the switch were adjusted for the moderate and severe visibility levels by adjusting the appropriate brightness and contrast controls to achieve the desired visibility. In these two positions the contrast was reduced while holding the brightness (luminance) level of the white calibration square at a constant 35 ft. lamberts. In this way the relative brightness of the display was held constant across viewing conditions. Once these controls were preset to achieve the desired viewing condition they were not changed. The cameras and lighting were checked both prior to and after testing to insure that the correct ratios were maintained.

The modulation contrast for the three visibility conditions was found by inserting the monitor screen luminance levels (β) into the following formula.

$$\text{Modulation Contrast (Percent)} = \frac{\beta max - \beta min}{\beta max + \beta min} \times 100$$

Visibility Condition	Modulation Contrast (Percent)
Clear	$\frac{35-1}{35+1} \times 100 = 94\%$
Moderate	$\frac{35-23}{35+23} \times 100 = 21\%$
Severe	$\frac{35-31}{35+31} \times 100 = 6\%$

Stereo Presentation

Perception of three-dimensional space occurs when the observer's left and right eyes are allowed to see the separate perspective views of an object. There are many techniques available today which allow the TV viewer to merge these two scenes into a single percept of 3–D space, i.e., refracting or reflecting stereoscopes, electronic or mechanical shutters and color or polarized filters used in conjunction with a half silvered mirror.

In the early 1970's a joint effort between Honeywell and the Naval Undersea Center resulted in the development of the PLZT (lead lanthanum zirconate titanate) stereoscopic viewer. This viewer utilizes the electro-optic shutter effect of the PLZT ceramic and, as with all shutter-type stereoscopic devices, it operates on the principles of alternately blocking and unblocking the perspective view for each eye of the observed object. "For example, when used with 2:1 interlace CRT displays, the pair of PLZT stereoscopic viewer lenses functions as electronic shutters that are 180 degrees out of phase with 50-percent duty cycles. For each frame, the perspective view for one eye is seen during the first field scan, while the other eye's view is blocked. This process is reversed for the second field scan to accommodate the perspective view for the other eye. Repetition of this sequence at normal television frame rates causes the observer to merge the time-sequenced perspective views for both eyes into a single image with a well defined depth of field." Reese and Khalafalla (1975)

LABORATORY EXPERIMENTS

In the following section, we turn to a series of laboratory experiments conducted to assess operator performance in a variety of TV-displayed task and visibility conditions. In all experiments, the major interest was in comparing performance using mono and stereo TV.

The first two experiments employed a category 1 task (Peg-in-hole). In addition to display and visibility parameters, we were interested in assessing differences in learning associated with instructional set, previous skill, and practice effects.

The third experiment employed a category 2 task (MLF) in a comparable experimental design as used for experiment 2, so that perceptual task and learning factors might be meaningfully evaluated.

A fourth experiment was neither designed nor conducted but logically follows from our discussions of visibility, task and learning factors. This experiment would be one designed to assess scene interpretation, possibly extracting or identifying an object from a highly unstructured, ambiguous and complex visual scene. The cow picture presented earlier is an example of scene interpretation. It is obvious at this point that designing such an experiment would challenge the best research minds involved in perception issues.

Peg-Task Experiments

The peg-task was chosen to represent that type of remote operator task which has abundant and relevant monocular cues in order to provide for both the recognition of objects, and their location in space. Other tasks in this category include drilling, tapping, threading, coupling, connecting, etc. They have in common the requirement for sensing the orientation of two pieces so that they can be properly aligned prior to engagement (which may include holding an alignment while imparting a rotation to the object).

The test operator's task in both studies was to position the manipulator arm to pick up one of the pegs from the starting block at the right front of the taskboard, grasp the peg firmly with the aid of flat sides cut into the peg, move the peg to one of the receiving blocks and insert it, then return to pick up the second peg and place it in the hole in the second block. In the first experiment, only time was scored, while in the second experiment, time and errors were both measured.

Experiment 1: Practiced Subjects. In experiment one (time-only), subjects were told to perform the peg-task as rapidly as possible, without regard to errors of mis-reach or mis-alignment.

In this first study, we attempted to reduce visual and motor learning and learning-to-learn effects to an absolute minimum. Subjects were given extensive training using direct and TV views of the taskboard, and included detailed coaching and verbal rewards for rapid performance. Thus, subjects were near their peak performance levels under ideal conditions when the study was begun. All subjects were run under all conditions in order to utilize the high reliability obtained in repeated design studies. The order of visibility conditions was from clear to moderate to severe to ensure that if any visual learning effects were still operative, they would accumulate over trials to the advantage of the moderate and severe visibility conditions. The highly structured taskboard provided vivid mono cues to form, texture, and location in depth.

The peg taskboard for both experiments is shown in Figure 7, with the manipulator extracting the second peg from the starting block. In order to ensure that the task would be visually guided on each two-peg trial, the taskboard was constructed so that it could be set to any of six positions at 15-degree increments of rotation, and to any of five elevations from flat to vertical.

The combined effects of six rotation positions and five elevation positions created a new alignment angle problem for 30 unique trials.

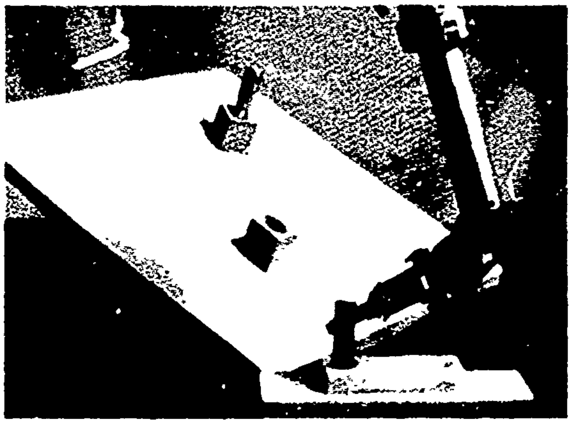


Figure 7. Peg-task in typical board position with good visibility (lower) and simulated underwater turbidity causing washout of contrast due to backscatter. The actual TV display is photographed to show the stimulus pattern presented to the subject.

The cameras aimed down at the taskboard start block from five feet, with a depression angle from the horizontal of 12 degrees. The position of the six rotation settings and five elevations relative to the cameras is shown in Figure 8. The receiving blocks were 2 x 2 x 2 inches, with an oversized one-inch hole for receiving the one-inch-diameter, four-inch-long peg. Tolerances were generous enough to permit a test subject to insert both pegs in only four seconds using direct viewing and his own hands. The fastest times using stereo television and the manipulator were on the order of eight seconds, a limit due to the inertia of the masterslave arm.

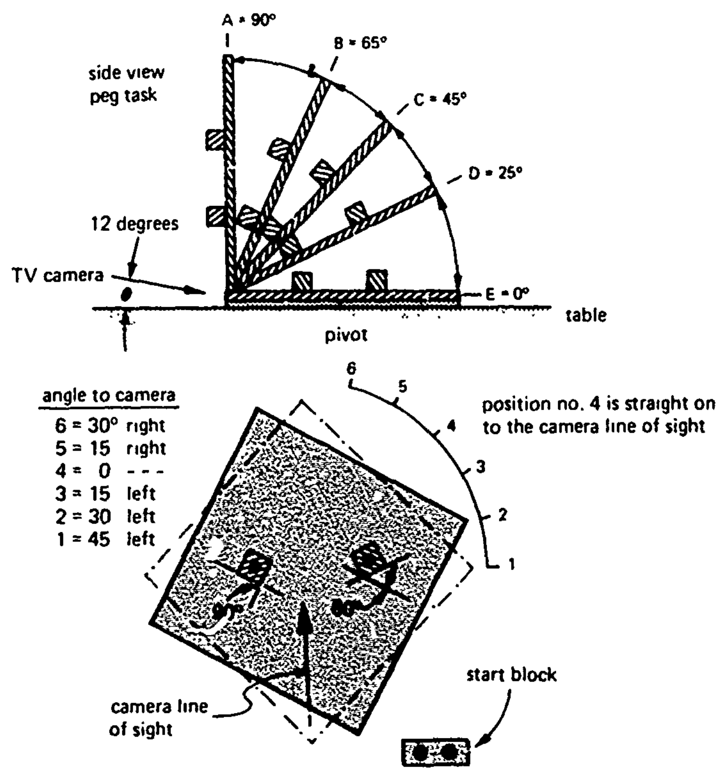


Figure 8. Elevation and rotation combinations for the peg-task. The upper figure is a view from the right side of the taskboard, while the lower figure is a top view with the board in elevation E (flat). The receiving blocks are shown with board in position 2, 30 degrees left.

The combination of camera and monitor characteristics resulted in an image of the receiving block and taskboard which was approximately four-tenths actual size, as the 24x 24-inch taskboard was approximately 10 inches on the display screen (a 17-inch Conrac monitor). The entire taskboard was painted with flat gray paint and coated with light gray flocking material to further reduce reflection and to simulate underwater sediment. The monitor and subject position were in a room next to the taskboard, as shown in Figure 9.

Figure 9. Test subject position with manipulator and TV. A single 17-inch monitor presented both stereo (via electronic shutter glasses) and mono.

Before discussing the detailed analysis of the experiment and the visual cues available under the stereo and mono TV conditions and under the three visibility conditions, there are additional photographs of the screen images as presented to the subjects. These will be presented now (Figures 10–14) so they will have been introduced for reference during the following discussion of cues.

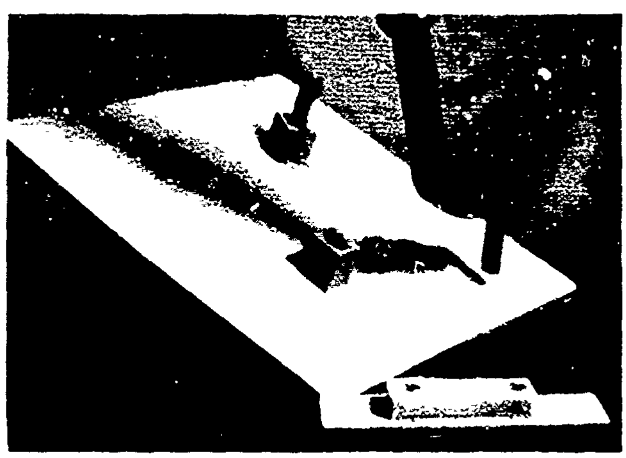


Figure 10. Example of the monocular cue often used by manipulator operators to determine how close they are to a surface. Since a shadow and object always converge with approach to a surface, well-placed lighting can provide very potent cues to final closure between manipulator and an object or surface. Projected shadows of two objects can be used to place the two objects in the same plane above a surface, and thus cause them to interact. As can be seen in Figure 11, such shadow cues can be lost with poor visibility, due to backscattering particulate matter in water.

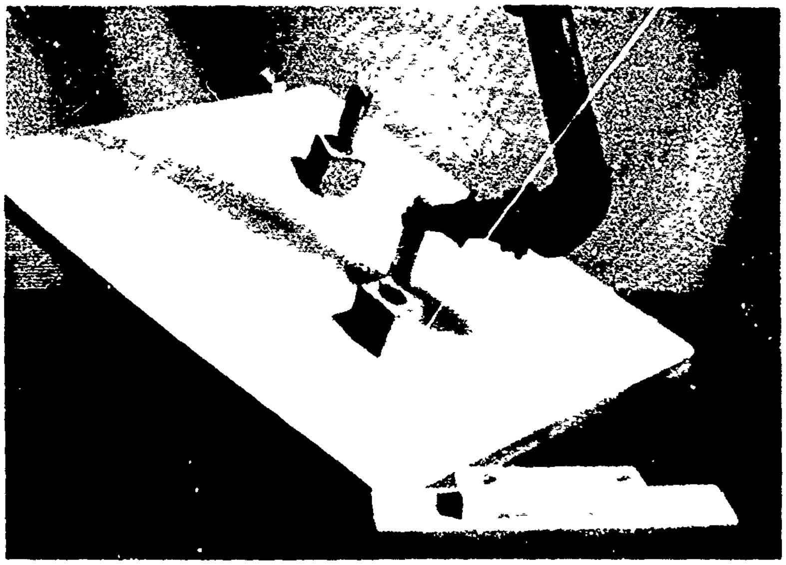


Figure 11. Example of interposition as a cue to rank-order of distance from the camera. In the clear view, note that the peg is seen to be obscured by the top surface of the block; even though it is properly aligned for insertion, it is not correctly positioned over the block center. Note the loss of this cue in the low contrast scene.

Figure 12. Closer view of TV display screen, showing some of the perspective cues to alignment. The peg axis (dotted line) must be parallel to the cube edge (solid line); the orientation and dimensions of the ellipses formed by the hole and the end of the peg must match. Note the peg's shadow (left arrow), and the sharply defined edge of the blo k formed by differential lighting (right arrow).

Figure 13. TV monitor screen in stereo mode (left photograph) and in mono (right photograph). The stereo system in use for this experiment used the odd-numbered lines for one eye's image, and the even-numbered lines for the other eye's image. This had the effect of reducing vertical resolution and contrast for light-colored objects surrounded by dark areas. When the subject was wearing the shutter glasses the left eye could see only the left block, the right eye only the right block (in the left photo). Although the peg looks lined up over the hole, it is immediately obvious in stereo that it is several inches behind the block.

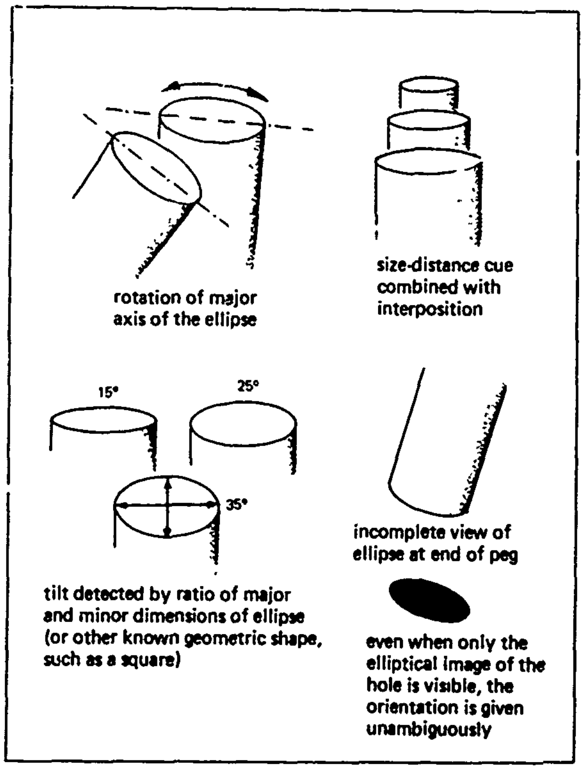
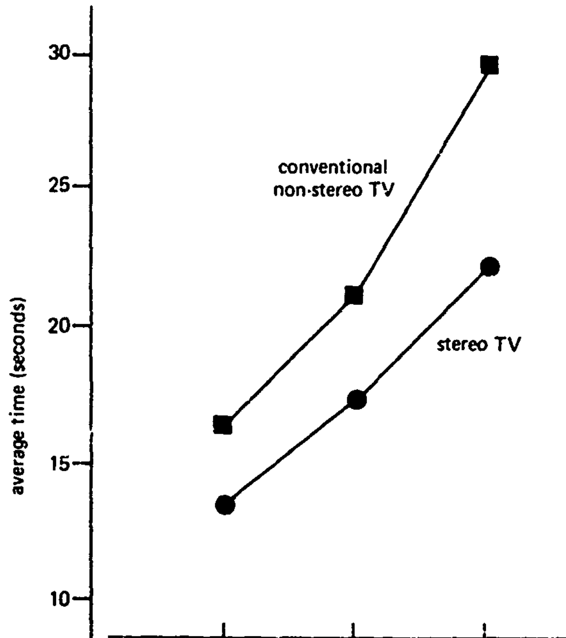


Figure 14. Some examples of geometric perspective cues to orientation which can be used for lining up tasks of this type. There must be enough effective resolution in the system to make use of such details. Under poor visibility conditions the ellipse was hard to see.

The first experiment employed six trained subjects who were highly skilled in the use of the CRL manipulator and in specific strategies for completing the peg-task in minimum times (errors not counted). In this experiment, learning effects were biased in favor of mono TV by performing the tasks in stereo (10 different positions) and then immediately repeating those same 10 positions in mono, for each visibility condition (clear, moderate, severe). Thus, any decrease in task time for stereo was in spite of a learning advantage gained during mono. Similarly, clear visibility conditions came first, followed by the moderate and severe conditions, so that any impairment of performance due to poor visibility would occur in spite of a learning advantage previously gained during clearer visibility conditions.

<u>Results and Discussion</u>. Figure 15 graphically shows the average peg-task performance times using stereo and mono TV for the three levels of visibility. Although, in this experiment, stereo showed a significant advantage over mono TV in terms of the ratio of task times, the absolute difference was on the order of 10 seconds at the most, and as little as three seconds in the clear visibility condition. While the difference in times between mono and stereo do not appear to be very large it must be remembered that the task is fairly simple and was performed by highly skilled operators. Any performance advantage must be multiplied by the number of times an operator would do a simple alignment movement during a complex manipulative task. It must also be remembered that the experimental design of this task was weighted against stereo and poor visibility and in all the individual subject averages, not one stereo score was worse than the corresponding mono score.



	clear $\bar{x}$	clear sd	moderate $\bar{x}$	moderate sd	severe $\bar{x}$	severe sd
mono	16.6	5.8	21.2	10.7	29.9	16.0
stereo	13.8	3.8	17.6	6.4	22.4	12.6

visibility

Figure 15. Average peg-task performance times for six practiced subjects using stereo and mono TV under three levels of visibility.

A totally Within Group's analysis of variance of the logarithmically transformed time scores is presented in Table 2 (See Winer, 1971, for a description of this model). It can be seen that the main effect of mono-stereo was highly significant ($F=27.0$ $p<.0025$), as was the effect of the visibility conditions ($F=21.88$ $p<.001$). Note that the interaction of mono-stereo by various levels of visibility was not significant. This may indicate that the loss in performance associated with decreased visibility occurred equally for both mono and stereo displays.

Table 2. Analysis of Variance Practiced Peg-Task.
(Log Time)

Treatment	df	MS	F	P<
Mono-Stereo (A)	1	0.081	27.0	.0025
Visibility (B)	2	0.175	21.88	.001
Subjects (S)	5	0.046	----	
AxB	2	0.002	1.00	N.S.
AxS	5	0.002	1.00	N.S.
BxS	10	0.008	4.00	.05
AxBxS	10	0.002		
Total	35			

This first study used highly practiced subjects because a major emphasis was made to select procedures that minimized potential learning effects and employed a statistical design to control for the variable effects contributed by individual subject performance. The data from this study and the low variability contributed by experimental error variance resulted in a demonstration that, even with the limited visual cue differences between the mono and stereo displays, stereo performance was superior to mono under all levels of visibility. Although the peg-task was not thought to be one where stereo would be critically important (it was chosen and designed to have strong mono cues), stereo nevertheless was able to cut performance times by 17 percent in clear and moderate visibility, and cut time by 25 percent in the severe visibility condition (see Figure 15). This result is similar to the 20 percent mono-stereo difference reported by Chubb (1964) using the same type task and manipulator under direct-viewed conditions.

Under these peg-task conditions, there are many useful and effective mono cues, as illustrated in Figures 10–14. Note also that since only time was being recorded, operators developed strategies with the force-feedback manipulator that maximized the amount of tactile feedback used to slip the peg into its final alignment. If the task had been drilling where there was no existing hole to provide tactile aid in alignment perpendicular to the surface, stereo would have been even more helpful.

The following is a general description of how subjects approached the peg-task and several of the skills and techniques which they developed to improve performance.

The first step in any remote work task is to interpret the scene so as to decide how to approach the problem. The overall view of the taskboard, as in Figure 7, provided strong cues to the tilt and orientation of the receiving blocks. Visual guidance was needed to grasp the peg with the jaws and for transport from the start block to the receiving block. Subjects rapidly learned ways to use tactile feedback in grasping the peg and in driving along a line of sight until the peg made contact with the hole and could be tipped in. Because there was no hesitancy or tendency to stop too soon, stereo made the travel time from start block to receiving block faster. The tendency to stop too soon was observed repeatedly under mono, probably because of depth uncertainties in the mono display. If the cameras had been closer to the task so that the critical features (ellipse axis of peg and block) were more finely resolved, it might have reduced performance time. The fact that additional resolution might help performance is supported by the ease with which operators could replace pegs in the plastic starting block if permitted to look directly through a small viewing port which could be opened in the wall.

In the severe visibility conditions, subjects were sometimes unable to see the starting position of the pegs and were barely able to make out where the receiving blocks were by the faint dark spot of the hole. In some stereo trials under the most sévere visibility, it was observed that the stereo system in use then (see Figure 13) was actually reducing contrast just below threshold, whereas in mono it was still visible. It should be noted that time performance was still better with the stereo system despite its reduced resolution, reduced contrast for light objects, and the more bothersome visual noise (in a spatial frequency sense) due to a raster pattern twice as coarse as in mono. The raster became an even greater problem when subjects would lean closer to the monitor in order to reach a difficult position with the manipulator. The visual interference caused by a high-contrast line pattern is demonstrated in Figure 16. By holding the page much farther away, or by moving it gently up and down, the eye can be clearly seen and even the eyelashes come into view.

Another factor which could have reduced stereo advantage is the tendency to tilt the head when reaching for difficult spots with the manipulator. When the subject's eyebase was no longer parallel with the eyebase on the TV screen, vertical disparity could cause loss of stereo fusion.

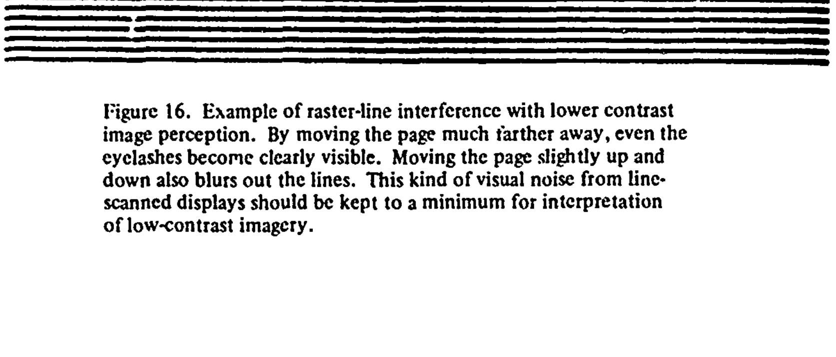


Figure 16. Example of raster-line interference with lower contrast image perception. By moving the page much farther away, even the eyelashes become clearly visible. Moving the page slightly up and down also blurs out the lines. This kind of visual noise from line-scanned displays should be kept to a minimum for interpretation of low-contrast imagery.

The visual cues used by the operators in mono and stereo were (1) simple contrast of the hole against the washed-out scene in severe visibility; (2) line-of-sight manipulator techniques when in mono, or in poor visibility, stereo; (3) strong perspective cues showing the taskboard alignment and orientation; (4) the linear alignment cues from the receiving block edges and ellipses; (5) shadow cues to assist in seeing when the peg/manipulator was going to touch the surface of the taskboard; and (6) interposition cues to tell when the peg was being placed behind or in front of the desired position (as in Figure 11). Tactile feedback was utilized more in the poorer visibility conditions and for the tipping in of the final alignment.

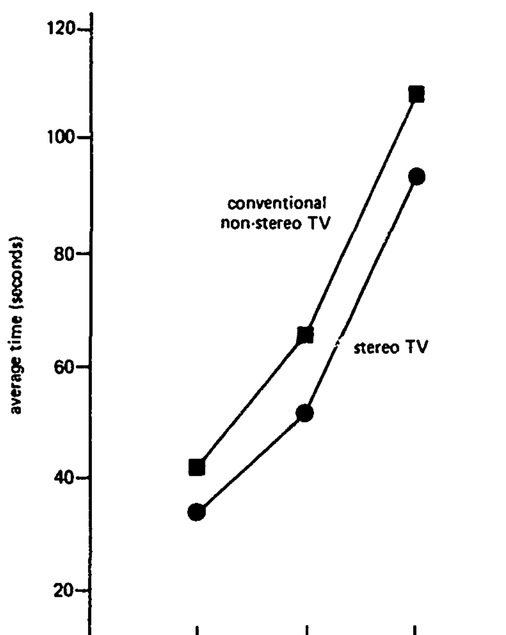
In stereo, the above cues were available, plus the strong and unambiguous primary cue to distance derived from horizontal disparity between left and right retinal images. This additional spatial information, in addition to providing a clear, dimensional image, permitted the subjects to move from the start block to the receiving block more confidently and quickly.

Experiment 2: Unpracticed Subjects. The second study was also designed to assess the effects of mono versus stereo views for three conditions of visibility and under conditions where subjects were not familiar with the task or taskboard, but had experience with remote viewing and manipulation. This study employed the same taskboard that is rich in monocular cues, consists of familiar shapes and forms, and requires relatively little discrimination of depth. Subjects were 16 NOSC employees, all of whom had some previous remote manipulator experience but were naive to this task. Subjects were randomly assigned to either mono or stereo viewing conditions and all were given 10 trials each under severe, moderate, and clear visibility conditions in that order. Subjects were given limited practice in removing the peg from the starting block under all these conditions. They were instructed to position the pegs in their respective holes, being very careful not to drop the peg or make unnecessary contact with the taskboard, as both time and error performance was recorded. The taskboard's rotation and elevation position was changed for each trial in order to reduce body position learning and to maximize reliance on visual cues. The purpose of ordering the visibility conditions from severe to moderate to clear was to minimize the carryover of visual information from one condition to the next. Previous research (Merritt, 1978) has shown that the information available in even one clear look at the taskboard scene can be utilized by the operator in later trials under reduced visibility. This same reasoning led us to use a Between-Group design for the mono versus stereo condition. Thus, for any subject in the mono condition, the only visual information that could carry over from trial to trial would be the mono cues present in that visibility condition and those from any preceding visibility conditions. While the carryover of visual information across visibility conditions was minimized by severe-moderate-clear order, any improvement in performance that might result from practice would be in favor of the clearer visibility conditions.

The expected consequence of these procedures was to reduce all non-visual factors to an absolute minimum and to maximize monocular depth cues in both the mono and stereo displays. Whereas these procedures are very unlike those encountered by the remote vehicle operator in real life situations, they provide an adequate and necessary test of the "pure" effects of the independent variables. It is our belief that all programatic research on remote operator performance should begin with such an assessment of the "pure laboratory" effects of the many variables associated with remote operator performance. It is only then that a meaningful assessment can be undertaken of the conditions which are imposed by the more real-life circumstances faced by the remote operator; that is, the effects of experience that are

a result of the uniqueness of each set of underwater task visibility conditions and target unfamiliarity.

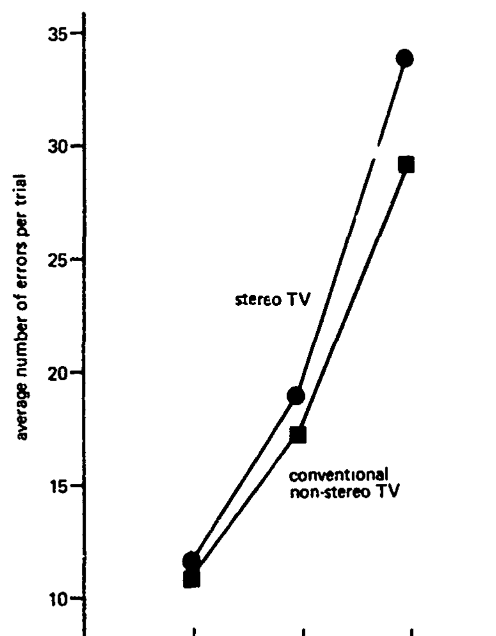
Results and Conclusions. The results of the second peg-task are presented graphically in Figures 17 and 18. Also, a Three-Way Between-Groups Analysis of Variance was employed which had mono-stereo (A) conditions as the Between-Groups's main effect, and visibility (B) and trials (C) as the Within-Group's main effects (see Winer, 1971). The trial main effect was a nuisance variable employed to account for variance associated with the task and trial order effects, which would normally be pooled with the experimental error variance term. The reduction of experimental error variance thus increases the overall sensitivity of the statistical analysis. Tables 3 and 4 present the three-way analysis of variance for the log transformed time and error scores. It can be seen that the main effect of visibility is highly significant ($F=101.55, P<.001$) for both time and ($F=8$[illegible]$.46, P<.001$) for errors. It can also be seen that the mono-stereo differences are not statistically significant for time or errors, even though stereo is consistently better on all points plotted for time in Figure 17. As Figure 18 shows, stereo performance is slightly worse when error scores are plotted. Note also that the visibility by mono-stereo interaction is not statisitically significant. This indicates that the degree of decrement associated with the visibility levels was similar in the mono and stereo conditions.



	clear $\bar{x}$	clear sd	moderate $\bar{x}$	moderate sd	severe $\bar{x}$	severe sd
mono	42.24	6.4	65.36	14.3	108.14	30.5
stereo	34.27	4.9	52.80	13.9	94.84	20.5

visibility

Figure 17. Average peg-task performance times. Note that while stereo results in consistently better performance, these differences are not statistically significant.



	clear $\bar{x}$	clear sd	moderate $\bar{x}$	moderate sd	severe $\bar{x}$	severe sd
mono	11.1	1.6	17.9	4.1	29.0	6.7
stereo	11.7	3.2	18.8	4.2	34.2	6.3

visibility

Figure 18. Average peg-task error performance.

Table 3. Analysis of Variance Unpracticed Peg-Task
(Log Time)

Source of Variation	df	MS	F	P<
Between-Groups	15			
A (Mono-Stereo)	1	0.855	1.8505	N.S.
Subjects W. Gr.	14	0.4605		
Within-Groups	464			
B (Visibility)	2	5.9902	101.5533	.001
AxB	2	0.0154	0.2608	N.S.
BxS	28	0.0590		
C (Task-Trials)	9	0.2134	10.1748	.001
AxC	9	0.0318	1.5176	N.S.
CxS	126	0.0210		
BxC	18	0.0629	3.0020	.001
AxBxC	18	0.0250	1.1934	N.S.
BCxS	252	0.0210		
Total	479			

Table 4. Analysis of Variance Unpracticed Peg-Task
(Errors)

Source of Variation	df	MS	F	P<
Between-Groups	15			
A (Mono-Stereo)	1	.449	.819	N.S.
Subjects W. Gr.	14	.549		
Within-Groups	464	6.9609	89.46	.001
B (Visibility)	2	.0806	1.036	N.S.
AxB	2	.0778		
BxS	28	.1404		
C (Task-Trials)	9	.1404	4.2644	.001
AxC	9	.0326	.9901	N.S.
CxS	126	.0329		
BxC	18	.0558	1.6027	.05
AxBxC	18	.0401	1.1516	N.S.
BCxS	252	.0348		
Total	479			

It is likely that the lack of sensitivity involved in the Between-Group's design contributed by the high degree of inter-subject variability in performance across all trials is responsible in part for the lack of a significant mono-stereo difference. It might be pointed out here that while our subjects were informed that errors were to be avoided, time-error tradeoffs occurred between the subjects which were uncontrolled; that is, each subject employed an individually determined tradeoff criterion during the performance of the tasks. It is also possible that inexperienced subjects were just unable to utilize the limited perceptual cue of binocular disparity, and it is also possible that the subjects learn (or utilize the visual cue information) differently in the mono and stereo conditions. In order to evaluate this second possibility, a split-half trials analysis of the three ten-trial blocks was conducted. The results indicate that there was no differential learning in the severe and moderate trial blocks, but markedly different learning under clear conditions, with the mono group showing a major learning effect ($\underline{t}$=4.52,$p<.005$), the stereo group showing no clear condition learning effect ($\underline{t}$=1.81,$p<.10$). It is probable that this differential learning is partially responsible for the lack of mono-stereo performance; however, it is felt that since this task was chosen and designed to have very strong mono cues, the results that show little or no improvement in stereo performance are not unexpected.

It will be recalled that our testing procedures were designed to reduce the impact of non-visual factors and to maximize monocular depth cues. The results of this study indicate that we were successful in this regard. They further indicate support for the argument that our subjects were better able to utilize the monocular cues under the clear condition and that stereo cues remain relatively stable under different levels of visibility.

Messenger-Line-Feeding (MLF) Task Experiment

Experiment 3. Subjects were 20 NOSC employees assigned randomly to mono or stereo conditions. They all had previous remote manipulator experience but were naive to this particular task.

The messenger-line-feeding (MLF) type of task was designed to represent a class of tasks such as line attachment, sample gathering, and certain simple salvage tasks (Category 2 in preceding discussions).

The MLF task (see Figure 19) duplicates the condition of unfamiliarity which often makes TV imagery so difficult to interpret in the reduced-cue situation found in the ocean environment. The taskboard surface is irregularly shaped with a plaster-like material in which the hoops are embedded. The irregular shape, as contrasted with the clean, flat taskboard used for the peg-task, is a representation of the way marine growth and corrosion can alter the contours of objects on the seafloor. The task is modeled after an actual operation in which a remotely-manned tether vehicle recovered an anchor chain at a depth of 600 ft. The hoops present the same appearance as the semi-buried links of that anchor chain, through which a hoisting line had to be threaded. The taskboard, three feet square, holds four 18 by 18 inch sections fitted with hoops (akin to a croquet wicket in size, made from various diameters of tubing), three to five hoops per quadrant.

The board was rotated to a new position each trial, so the terrain could not be learned. The number and type of layout precluded even the experimenters from learning the spatial positions.

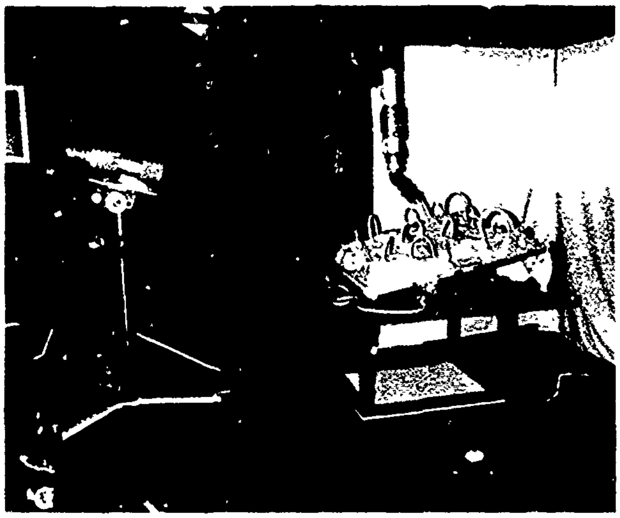


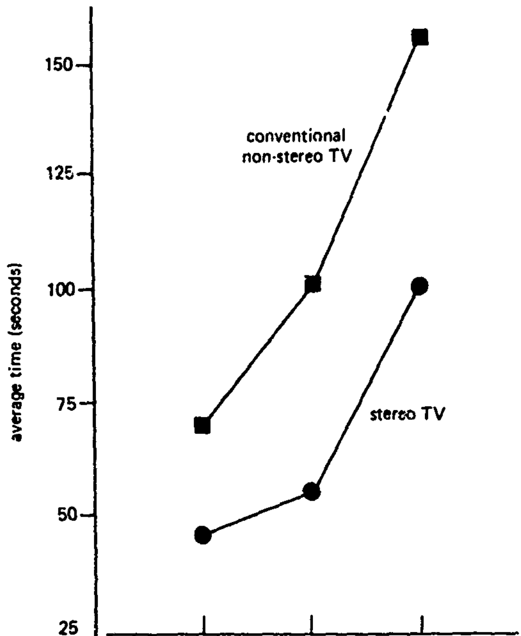
Figure 19. Overall view of the cameras and MLF taskboard, which could be rotated to 24 15-degree position increments around a full circle. The board was inclined 14 degrees, the cameras down 14 degrees, so the observation angle varied from 28 to 0 degrees. The bottom view is a photograph, not the TV display.

Subjects were informed that time and error scores were being recorded so that while speed was an important part of their performance, accuracy was also important.

The task consisted of threading a half-inch rope through two hoops as designated by the experimenter just prior to starting. The subjects were not shown the board before the tests, and were given practice trials using an older prototype of this taskboard immediately prior to the experiment. Subjects in each group were then given 10 trials each under severe, moderate, and clear visibility conditions.

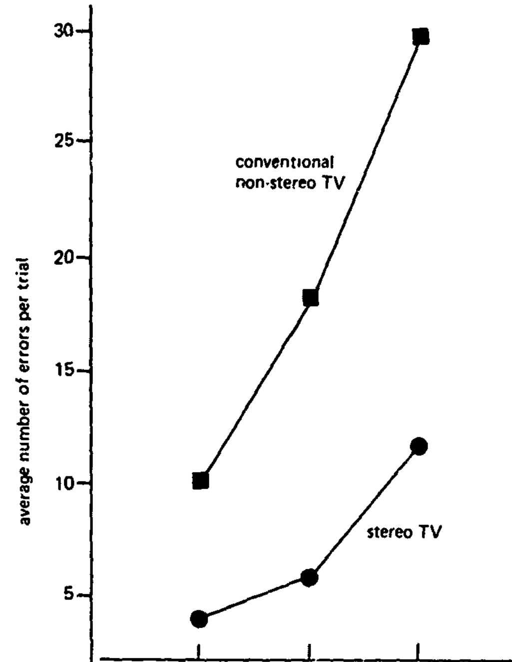
Results and Discussions. The results of the MLF experiment are presented graphically in Figures 20 and 21.

The results showed that subjects took 50 percent longer to do the task in mono, and had over twice the number of errors (inadvertent contacts with the board) than in stereo. Note that we employed the same Between-Groups design, so there was no learning advantage for either mono or stereo.



	clear $\bar{x}$	clear sd	moderate $\bar{x}$	moderate sd	severe $\bar{x}$	severe sd
mono	66.4	30.4	98.8	34.5	153.7	43.5
stereo	47.4	20.2	56.2	18.7	98.3	60.2

Figure 20. Average Messenger-Line Feeding (MLF) task performance times.



	clear $\bar{x}$	clear sd	moderate $\bar{x}$	moderate sd	severe $\bar{x}$	severe sd
mono	10.0	4.8	18.9	8.1	29.6	9.0
stereo	4.5	1.7	6.8	2.9	12.5	8.5

Figure 21. Average Messenger-Line Feeding (MLF) task error performance.

Table 5 presents the results of the three-way Between Groups Analysis of Variance for the lo[illegible]e time scores. It can be seen that the main effects of mono-stereo (A) and visibility [illegible]ditions (B) were highly significant ($A^*F=14.36, p<.0025$; $B^*F=25.45, p<.001$). Additionally, the A by B interaction was also significant (A by $B^*F=4.88, P<.05$). A corresponding analysis was completed on the error scores, and in all cases the results were identical to those obtained for the time analysis. These data are presented in Table 6.

As can be seen in Figure 19, the hoops were painted with a light gray flocking material to add a fuzzy surface, not unlike that found on undersea objects. The hoops, then, were already low in contrast, and when the moderate and severe visibility conditions were imposed, the cue of interposition was degraded below effective threshold. The lighting was from several different angles and from large-area sources so as to duplicate somewhat the diffuse lighting found in the sea. The irregular shapes did not provide linear perspective cues, nor did it

Table 5. Analysis of Variance of MLF Task.
(Log Time)

Source	df	MS	F	P
Between-Groups				
(A) Mono-Stereo	1	5.72	14.01	.0025
Sub Within gr	18	.4085		
Within-Groups				
(B) Visibility	2	5.316	110.54	.001
AxB	2	.1659	3.45	.05
BxS	36	.048		
(C) Task Trials	9	.323	11.35	
AxC	9	.043	1.52	
CxS within gr	18	.598	19.55	.001
AxBxC	18	.075	2.46	.01
BCxS within gr	252			
TOTAL	479			

Table 6. Analysis of Variance of MLF Task
(Errors)

Source of Variation	df	MS	F	N
Between-Groups	19			
A(Mono-Stereo)	1	20103.595	54.0120	.001
Subjects W. Gr.	18	372.2048		
Within-Groups	580			
B(Visibility)	2	9557.9687	84.1803	.001
AxB	2	1727.6958	15.2164	.001
BxS	36	113.5417		
C(Task-Trials)	9	555.6943	9.8182	.001
AxC	9	74.4705	1.3158	N.S.
CxS	162	56.5984		
BxC	18	786.1133	11.9892	.001
AxBxC	18	155.9180	2.3779	.01
BCxS	324	65.5687		
Total	599			

provide cues derived from known sizes and shapes; the hoops were made from tubing of various diameters, ranging from 3/8th inch to over 1 inch diameter. In Figure 19, the confounding of size and distance can be seen, where a small diameter hoop and a large diameter hoop are side-by-side in the foreground, bottom. It should be noted again that the photograph at the bottom in Figure 19 is not (as in the previous peg-task photography) a picture of the TV display; it is a close-up photograph taken directly with a 35mm camera, and shows more detail than was available to the test subject under the severe visibility condition. Due to the confounding of monocular size/distance cues, the board was very difficult to perceive in mono, but in stereo the whole arrangement of the elements was immediately clear. The only impediment for those operating in stereo was to learn skill and techniques with the manipulator for this task.

The time and error performance scores shown in Figures 20 and 21 indicate a significant advantage for stereo TV in this type of remote manipulator task, due to the reduced level of monocular cues available. Unlike the peg-task, the MLF task was designed to control or eliminate many of the cues which are often present in simple laboratory tasks used to evaluate manipulator variables without regard to the visual display variables. The interaction of the visibility factor with the stereo-mono factor shows that stereo TV is degraded less by poor visibility than is mono TV, with the same general curve-shape for both time and error scores. Under the severe visibility condition, the lower resolution and contrast available in the stereo system tended to work against the stereo advantage (see Figure 13) as shown in the increased slope of the stereo curve (Figures 20 and 21). Even with these disadvantages (which were due only to the type of stereo system employed at the time, rather than to stereo systems in general) stereo performance times were significantly better than those with mono TV, and error scores were greatly reduced. The importance of the error scores can be placed in perspective by considering the critical nature of tasks such as munitions recovery, handling of radioactive objects lost at sea, dropping or breakage of expensive or irreplaceable tools or equipment, and so on.

The importance of the interaction of visibility with stereo TV points up the relative immunity of stereo systems to noise and contrast reduction, both of which are very common in the undersea imaging environment. It was in consideration of this characteristic advantage of stereo in photointerpretation which lead to one of our research hypotheses: that stereo would provide an increasingly significant advantage over mono as visibility conditions and task object complexity became more difficult. The results of our research confirm this hypothesis.

CONCLUSIONS AND RECOMMENDATIONS

This study sought to examine the relative performance advantage obtained in a manipulator task when the cue of binocular parallax is added to the usual televised scene.

As scene complexity and object ambiguity increased (our category 2 task), the advantage of a stereo display became more pronounced. We believe this to be due to several factors. First, as we have demonstrated, with decreased visibility, the cues to distance given monocularly are reduced proportionally. Binocular disparity is less sensitive to degradation; therefore, stereo performance remained consistently higher. Second, in complex, highly unstructured and uncertain visual scenes, the dimension of scene interpretation becomes an increasingly important factor. Binocular disparity provides significant information under these conditions.

Other types of perceptual information would also be expected to improve performance. Included in this latter type are motion parallax and color registration.

Motion parallax provides relative cues to distance by virtue of the information resulting from the observation of differential object-displacement when the head is translated on the frontal-parallel plane (X dimension). Objects at the fixation point appear stationary, while objects beyond this point move at rates which are dependent upon their distance from the fixation point.

In our stereo system, the movement of the objects are not faithfully reproduced. In fact, objects beyond the convergence point (the face of the display monitor) appear to move opposite to the expected direction. This is because the movement-compensation mechanism in the brain expects the object-images on the retina to be displaced, and therefore compensates for the head movement by interpreting movement in the opposite direction. We have termed this apparent movement "pseudo-parallax." It is unknown to what extent the inappropriate motions have contributed to errors in depth or distance perception. Additional sources of potential error may be contributed by the mismatch between accommodation and convergence. Previous research in perception indicates that the perceived absolute distance in mismatch circumstances results in a compromise between the two cues (Ono, Mitson and Seabrook, 1971).

The relationship between parallax cues given monocularly and binocularly, and the magnitude and direction of error introduced by "pseudo-parallax" cues can only be determined empirically. If we are to continue to utilize the advanced display systems to accomplish more and more sophisticated and hazardous missions, we must more fully understand the contribution of these variables to task performance.

Additional research needs to be addressed toward tasks involved in scene interpretation. Category 3 tasks need to be identified and performance measures obtained utilizing those visual cues which provide increased information for scene interpretation. Color is an extremely important cue, probably as important as binocular disparity under some scene conditions.

Other features found in direct vision need to be assessed to determine their utility, in addition to binocular and motion parallax cues. These include accommodation, convergence, color, improved resolution, improved gray-scale rendition, color rendition, and an integrated visual-motor space. The relative performance advantages of these visual-perceptual-motor features can only be determined through experimentation with generic tasks and an advanced manipulator system which does not constitute the major limiting factor to performance. Even when such advantageous features as color or stereo are not essential for task completion, due to the abundance of monocular cues and the adaptability of the operator, stereo will still reduce the time required for completion and will greatly reduce the number and severity of contact errors which could be critical in hazardous situations.

With the advent of advanced master-slave manipulators with excellent force-feedback (which are now available in hot-cell laboratories), and the utilization of improved display systems employing stereo, color, high resolution, motion parallax, etc., man's capabilities will soon be extended into depths and hostile environments which until now have not been possible.

REFERENCES

1. Biberman, Lucien M. (Ed.), Perception of Displayed Information. Plenum Press, New York. 1973.

2. Chubb, G. P., A Comparison of Performance in Operating the CkL-8 Master Slave Manipulator under Monocular and Binocular Viewing Conditions. Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio. MRLD-TDR-64-68 (AD 608791), October 1964.

3. Forgus, R. H., Perception. New York: McGraw-Hill, 1966.

4. Funk, G. S., S. B. Bryant and P. J. Heckman, Jr., Handbook of Underwater Imaging System Design, NUC TP 303, San Diego. July 1972.

5. Graham, C. H. (Ed.), Vision and Visual Perception. John Wiley and Sons, Inc., New York, 1965.

6. Gregory, R. L., Eye and Brain. New York: McGraw-Hill, 1966.

7. Gregory, R. L., The Intelligent Eye. New York: McGraw-Hill, 1970.

8. Helmholtz, H. V., Treatise on Physiological Optics. (Reprint of the translation of the 3rd Edition; Ed. J. P. C. Southall) Washington, D.C.: Optical Society of America, 1924. (Available as a Dover reprint, 1963.)

9. Hennessy, R. T., Personal communication, 1975.

10. Hochberg, J. *In* J. W. Kling & L. A. Rigs (Eds.), Woodworth & Schlosberg's Experimental Psychology (3rd ed.). New York: Holt, Rinehart & Winston. 1971.

11. Hudson, E. E. and G. Culpit. Stereo TV Enchancement Study. Final Technical Report. Kollsman Instrument Corp., Syosset, New York. February 1968.

12. Julesz, B., Foundations of Cyclopean Perception. Chicago: University of Chicago Press, 1971.

13. Kama, W. N. and R. C. DuMars, Remote Viewing: A Comparison of Direct Viewing, 2-D and 3-D Television. Report AMRL-TDR-64-15, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio. February 1964.

14. Merritt, J. O., Visual Information Requirements for Remotely-Manned Undersea Work Systems. TP-2708-2, Human Factors Research, Inc., Goleta, CA, September 1978.

15. Ogle, K. N. *In* H. Davson (Ed.), The Eye (Vol. 4). New York: Academic Press, 1962.

16. Ono, H., L. Mitson, and K. Seabrook. Changes in Convergence and Retinal Disparities as an Explanation for the Wallpaper Phenomenon. Journal of Experimental Psychology, 1971, **91**, 1-10.

17. Pepper, R. L. and R. E. Cole, Display System Variables Affecting Operator Performance in Undersea Vehicles and Work Systems. NOSC TR 269, Naval Ocean Systems Center, San Diego, June 1978.

18. Pepper, R. L., R. C. Cole, J. O. Merritt, and D. Smith. Operator Performance Using Conventional or Stereo Displays. Optical Engineering, 1978, 17(4), 411–415.

19. Pesch, A., Behavioral Cybernetics Theory Applied to Problems of Ship Control and Manipulator Operation in Small Submarines. Proceedings of the 2nd Marine Systems and ASW Conference. Los Angeles, California. August 1967.

20. Roese, J.A., A. S. Khalafalla, Stereoscopic Viewing with PLZT Ceramics, Ferroelectrics, Vol. 10, 1976.

21. Uhrich, R. and R. B. Fugitt. Investigation of Control and Viewing Techniques for Use With Remote Undersea Manipulator. NOSC TR 233, Naval Ocean Systems Center, San Diego. 1 May 1978.

22. Vlahos, P. The Three-dimensional display: Its cues and techniques. Information Display, 1965, Nov–Dec, 10–17.

23. Winer, B. J. Statistical Principles in Experimental Design, McGraw-Hill Book Co., New York, 1971.